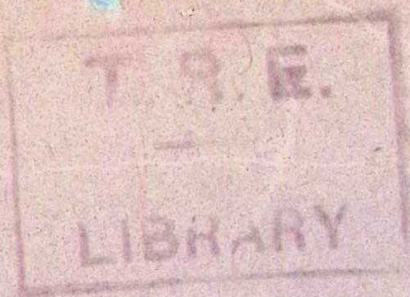


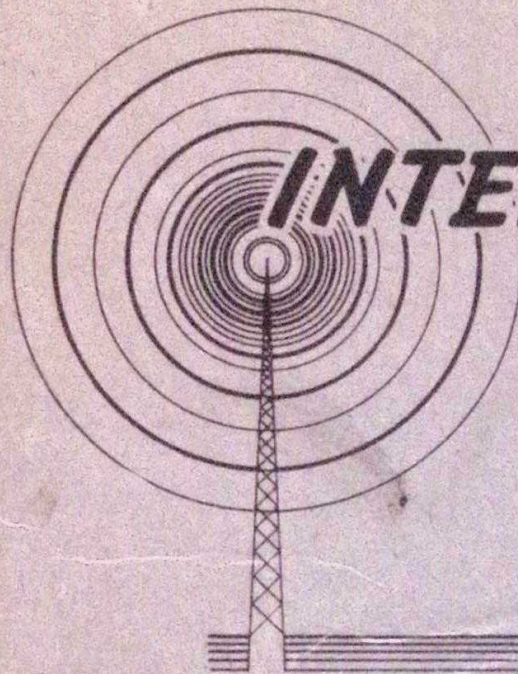
Copy 2

RESTRICTED



BR 1600A - Admiralty
CODE N°1563 - War Office
A.P. 1093E - Air Ministry

FIRST EDITION
JUNE 1946



INTERSERVICES RADAR MANUAL

VOLUME II
RADAR TECHNIQUES

*Prepared by direction of the
Minister of Supply*

J. S. Franks

*Promulgated by order of the
Air Council*

W. B. Brown

AIR MINISTRY

Admiralty No. BR 1600 A
War Office Code No. 1563
Air Ministry No. AP 1093 E

Volume II

NOTE TO READERS

Air Ministry Orders and Leaflets may affect the subject matter of this publication. Where possible, Amendment Lists are issued to bring this into line. When an Order or Leaflet is found to contradict any portion of this publication, the Order or Leaflet is to be taken as the over-riding authority.

When this volume is amended by the insertion of new leaves the new or amended technical information is indicated by a vertical line in the outer margin. This line is merely to denote a change and is not to be taken as a mark of emphasis.

Each such leaf is marked in the top left-hand corner of the right hand page with the number of the A.L. with which it was issued.

Foreword to the Restricted Edition

This book is the second volume of a three-volume project which comprises :

Volume I.	Radar Principles
Volume II.	Radar Techniques
Volume III.	Radar Systems and Performance.

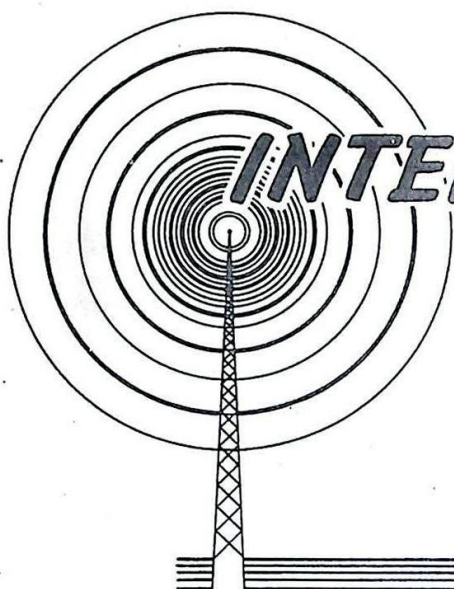
The objective of the authors and editors has been to produce an Inter-Service Radar Manual dealing with generalities (excluding any discussion of particular equipments, which is better left to Service Handbooks dealing with individual sets) and of such a standard as will permit its use for wide reading amongst personnel engaged in the employment, repair and maintenance of Radar equipment.

The need for an exposition of techniques having been deemed more urgent than the other matters to be dealt with in this composite work, the second volume has been completed and is now presented in advance of either of the other two; and although some work has been done in preparation of the latter, the disorganisation of the writing panel consequent upon the capitulation of the Axis powers has been such as to preclude any possibility of the early completion of the work.

In order to reduce the size of the book, relevant matter to hand in widely circulated Service handbooks has been excluded, and previous acquaintance of the reader with one or the other of these is therefore assumed, as is evident from the references in the text. Certain paragraphs of this volume are starred thus :- ~~xxx~~ . It is suggested that these paragraphs be omitted at a first reading or by readers who wish to avoid discussions which are largely of a mathematical nature.

The Restricted edition is now circulated to select groups of readers in advance of a more general (and possibly public) availability in finished form, with a view to collecting opinions and suggestions from those branches of the Services and other departments most likely to find use for the printed work. Communications regarding these matters would be welcomed at an early date, as the final printing cannot be unduly delayed, and should be addressed to

The Editor,
I.S.R.M.,
Radio Board,
143, Piccadilly,
London, W.1.



INTERSERVICES RADAR MANUAL

VOLUME II RADAR TECHNIQUES

LIST OF CONTENTS

CHAPTER 1

LINEAR CIRCUIT ANALYSIS

1. Introduction
2. Circuit Laws and Conventions
- Characteristics of Circuit Elements
 3. Passive Elements
 4. Active Elements
- Linear Circuits
 5. General
 6. Superposition Theorem
 7. Reciprocity Theorem
- Representation of Sinusoidal Currents and Voltages in Linear Circuits
 8. General
 9. Complex Quantities
 10. Vector Diagrams
- Impedances and Admittances
 11. General
 12. Helmholtz's, Thevenin's or Norton's Theorem
 13. Star-Delta Transformation
 14. Series Networks
 15. Parallel Networks
 16. Q- Factor of a Component
 17. Equivalent Series and Parallel Networks

18. Frequency Characteristics of Two-Terminal Networks
19. Resonance
20. Coupled Circuits
21. Broad-Banding

CHAPTER 2

RESPONSE OF LINEAR CIRCUIT ELEMENTS TO VOLTAGE PULSES

1. Introduction
 - Instantaneous Application of Changes of Voltage to Circuits Containing Capacitance and Resistance
 2. Instantaneous Application of a Change of Voltage to a Series C-R Circuit
 3. Application of Rectangular Pulse of Voltage to a Series C-R Circuit
 4. Application of a Succession of Rectangular Pulses of Voltage to a Series C-R Circuit
 5. Instantaneous Application of a Change of Voltage to Circuits Containing More than One Condenser or Resistor
 - Instantaneous Application of Changes of Voltage to Circuits Containing Inductance and Resistance
 6. Instantaneous Application of a Change of Voltage to a Series L-R Circuit
 7. Application of a Rectangular Pulse of Voltage to a Series L-R Circuit
 8. Application of a Succession of Rectangular Pulses of Voltage to a Series L-R Circuit
 - 9. Disadvantages of a Series L-R Circuit Compared with a Series C-R Circuit
 - Instantaneous Application of Changes of Voltage to Circuits Containing Inductance, Capacitance and Resistance
 10. Instantaneous Application of a Change of Voltage to an L-C-R Circuit
 11. Instantaneous Application of a Change of Voltage to a Delay Network
 - Non-Instantaneous Changes of Voltage
 12. General Discussion
 13. Input Voltage Increasing Linearly with Time
 14. Linear Rise of Voltage of Finite Duration
 15. Pulses Involving Linear Variations of Voltage
 16. Succession of Pulses
17. Differentiating and Integrating Circuits

CHAPTER 3

NETWORKS

1. Introduction
2. Measurement of Input and Output Impedances
3. Symmetrical T-Sections and π -Sections
4. Matching
5. Gain and Loss
6. Iterative Sections
7. Attenuators
- Filters
 8. Ideal Requirements
 9. Simple High-Pass and Low-Pass Filters

10. Simple Low-Pass Filter Formed of Purely Reactive Elements
 11. Graphical Representation of Properties of a Simple Low-Pass Filter T-Section
 12. Properties of Simple Low-Pass Filter Π -Section
 13. Properties of Simple High-Pass, Band-Pass and Band-Stop Filters
 14. Other Types of Filters
 15. Delay Networks
 16. Use of Delay Line for producing Rectangular Pulses
- Resolving Networks
17. General
 18. Resistive Resolver
 19. Inductive Resolver
- Continuously Variable Phase-Shifting Devices
20. General
 21. The Goniometer
 22. The Capacity Phase-Shifter
 23. The Potentiometer Phase-Shifting Network
 24. 0-180° Phase-Shifting Network

CHAPTER 4

TRANSMISSION LINES

Introduction

1. Circuit with Distributed Constants
2. Transmission Lines

Travelling Waves on Uniform Loss-Free Transmission Lines

3. Application of Circuit Laws to Transmission Lines
4. Interference
5. Waves of Various Shapes
6. Sinusoidal Waves
7. Electromagnetic Waves on Transmission Lines

Uniform Loss-Free Lines : Reflection at the Termination and its Effects

8. Reflection of a Rectangular Wave
9. Coefficient of Reflection
10. Line Fed by a DC Generator
11. Deduction of Line Characteristics
12. Discharge of Open-Circuited Line
13. Reflection of Sine Waves
14. Standing Waves on a Uniform Open-Circuited Loss-Free Line
15. Standing Waves on a Uniform Short-Circuited Loss-Free Line
16. Input Impedance of a Uniform Loss-Free Line for any Termination

Fundamental Line Characteristics

17. Distributed Elements for a Loss-Free Line
18. Distributed Elements for a Lossy Line
19. Characteristic Impedance
20. Propagation Constant
21. Vectorial Representation

Lines with Low Losses

22. Effects on Line Characteristics
23. Effects of Slight Losses on Travelling Waves
24. Effects of Slight Losses on Standing Waves
25. Energy Losses

Resonant Lines

26. General Nature of Resonance in Lines
27. Q-Factor of a Resonant Length of Line
28. Lecher Bars
29. Metallic Insulators
30. Quarter-Wave Sleeve Rejector (Rotating Joint)
31. Stub Reactances

Matching

32. Reasons for Matching
33. Half-Wave Transformer
34. Quarter-Wave Matching Sections
35. Double Quarter-Wave Line
36. Matching by Stubs
37. Slugs
38. Balance to Unbalance Transformer
39. Standing Wave Indication
40. Common T/R Circuits

Limitations of Transmission Lines

41. Resistive and Dielectric Losses
42. Resistive Losses
43. Dielectric Losses
44. Frequency Effects
45. Coaxial Cables with Low Losses
46. Other Losses in Open-Wire Feeders

Circle Diagrams

47. Introduction
48. Normalised Impedances and Admittances
49. The Cartesian Circle Diagram
50. Applications of the Circle Diagram
51. Numerical Examples
52. Use of Circle Diagram to Determine the Magnitude of the Load Impedance from a Knowledge of the Standing Wave Pattern
53. Application of Circle Diagrams to Lines with Low Losses
54. Application of Circle Diagrams to Matching Devices Constructed from Sections of Uniform Loss-Free Transmission Line

CHAPTER 5

WAVEGUIDES

Introduction

1. General
2. Properties of Electromagnetic Waves in Free Space
3. Behaviour of an Electromagnetic Field at the Surface of a Conductor
4. Waves Guided by Pairs of Conductors (Transmission Lines)
5. Limitations of Cables and Advantages of Waveguides at Centimetre Wavelengths
6. Waves in Waveguides Compared with those in Free Space and on Transmission Lines

Waves in a Rectangular Tube

7. Synthesis of an H-Wave
8. Pattern Velocity (Phase Velocity) of H-Waves
9. Propagation Between Parallel Plates and in Rectangular Tubes
10. The H_{01} -Wave in a Rectangular Guide
11. Cut-Off Wavelength of the H_{01} -Wave

12. Evanescent Modes
13. Cut-Off Wavelengths of Different Modes of H-Waves
14. Method of Launching an H_{01} -Wave in a Rectangular Waveguide
15. Wall Currents in the Case of the H_{01} -Wave
16. E_{nm} -Waves and H_{nm} -Waves

Waves in Circular Tubes

17. General
18. H_{01} -Wave in a Circular Guide
19. Classification of Waves in Circular Guides
20. The E_{01} -Wave in a Circular Guide
21. The Septate Guide
22. The H_{11} -Wave in a Circular Guide

Attenuation in Waveguides

23. Attenuation of Propagated Modes
24. Attenuation of Evanescent Modes

Cavity Resonators

25. General
26. Stationary Waves on a Waveguide
27. Field patterns in Cavity Resonators
28. Resonant Wavelength of the Cavity
29. Charges and Currents on Internal Surface of Resonator
30. Methods of Excitation of a Cavity Resonator
31. The Q-Factor of a Cavity
32. Applications of Cavity Resonators
 - (i) Wavemeters
 - (ii) Echo Boxes
 - (iii) Cavity Resonators in Centimetre Wave Oscillators

Choice of Waveguide Shape and Dimensions

33. General Considerations
34. Choice of Waveguide Dimensions
35. Geometrical Form of Waveguide Section

Standing Waves in Waveguides

36. Waveguide Impedances
37. Reflection from the Waveguide Termination
38. Practical Methods of Achieving a Reflectionless Termination
39. Attenuating Section
40. Reflection from an Irregularity in a Waveguide
41. Standing Wave Indicator
42. Elimination of a Reflected Wave
43. Matching Devices
 - The Capacitive Iris, (H_{01} -Mode)
 - The Inductive Iris, (H_{01} -Mode)
 - Inductive Wire (H_{01} -Mode)
 - Irises in Circular Waveguides (H_{11} -Mode)
 - Resonant Irises
 - The Q-Factor of Resonant Irises

44. Switches and Protective Devices
45. Reflectors in Waveguides

Reflections from Joints and Bends in Waveguides

46. General
47. Waveguide Joints
48. Bends in Waveguides
49. Rotary E-H Transformer

50. Common T/R with Waveguides

CHAPTER 6

ELECTRONIC DEVICES

Cathode Ray Tubes

1. Construction and Operation
2. Types of Tube
3. The Cathode
4. The Anode
5. The Control Electrode (Control Grid)
6. Brilliance
7. The Fluorescent Screen
8. Screens with Multiple Layers
9. The Skiatron or Dark-Trace Tube
10. The Glass-Envelope

Focusing Systems

11. General
12. Electrostatic Focusing
13. Magnetostatic Focusing
14. Gas Focusing in Soft Tubes

Deflection Systems

15. General
16. Electric Deflection
17. Magnetic Deflection

Distortions and Their Correction

18. Trapezium Distortion
19. Deflection Defocusing
20. Astigmatic Distortion
21. Bulb Charge and Deflector Plate Current
22. Stray Fields Leading to Distortion
23. Comparison of Focusing and Deflection Systems
24. Post-Deflector Acceleration
25. Double-Beam Tube

Power Supplies and Shift Networks

26. General
27. Power Supplies
28. Shift Voltage for Electric Deflection Systems
29. Shift Current for Magnetic Deflection Systems

Input Circuits

30. Input Circuits to Deflector Plates
31. Input Circuits to Grid and Cathode

Hard Valves

32. Introduction
33. Power Valves for Radar Transmitters
34. Suppressor-Grid Characteristics

Soft Valves

35. Summary of the Effect of Gas in Valves
36. Cold Cathode Diode
37. Hot Cathode Diode
38. Hot Cathode Triode
39. Spark Gaps
40. The Triggered Spark Gap
41. The Rotary Spark Gap
42. Soft Rhumbatron

Non-Ohmic Resistors

43. General
44. Metrosil
45. Thermistors

CHAPTER 7

VALVE AMPLIFIERS

1. Introduction
2. Types of Distortion
3. Conventional Symbols and Equivalent Circuits
4. Bias for Amplifiers
5. The Potential of the Screen Grid of a Pentode Used as an Amplifier
6. Anode Decoupling of Amplifiers
7. Miller Effect
8. RF Amplification
- Video-Frequency Amplification
 9. General Considerations
 10. The Amplification of an Instantaneous Change of Voltage
 11. The Amplification of a Rectangular Voltage Pulse
 12. The Frequency Spectrum Approach to Video-Frequency Amplification
 13. The Use of Pentodes in Video-Frequency Amplification
- Negative Feed-Back Amplification
 14. General Principles
 15. Voltage Feedback (Parallel Feedback)
 16. Current Feedback (Series Feedback)
 17. Bias for Feedback Amplifiers
- The Cathode Follower
 18. The Properties of a Cathode Follower
 19. Applications of a Cathode Follower
20. The Cathode Input (or Grounded Grid) Amplifier
- Paraphase Amplifiers
 21. General
 22. Single-Valve Paraphase Amplifiers
 23. Two-Valve Paraphase Amplifiers
- Limitations of the Use of Valves at High Frequencies
 24. General
 25. Transit Time Effects
 26. Circuit Limitations

CHAPTER 8

VALVE OSCILLATORS

Introduction

1. Uses of Valve Oscillators
 2. Fundamental Requirements of a Valve Oscillator
 3. Condition for Maintenance of Oscillations
- #### Oscillators Employing Mutual Inductive Coupling
4. Tuned Grid Oscillator
 5. Tuned Anode Oscillator
 7. Magnetostriction Oscillator
- #### Oscillators Employing Direct Coupling with a Single Tuned Circuit
8. Hartley Oscillator
 9. Colpitts Oscillator
 10. Dynatron Oscillator
 11. Transitron Oscillator
- #### Oscillators Employing Direct Coupling with Twin Tuned Circuits
12. General
 13. Common-Cathode Oscillator
 14. Common-Anode Oscillator
 15. Common-Grid Oscillator
 16. Earthing the Twin-Circuit Oscillators
 17. Types of Tuned Circuits Employed
 18. Crystal-Controlled Oscillators

19. Push-Pull Circuits
20. Resistance-Capacitance Oscillators
- Velocity Modulated Oscillators
 21. General
 22. Velocity Modulated Amplifier
 23. Velocity Modulated Oscillators in Common Use

Heil Tube
 Double Rhumbatron Klystron
 Sutton Tube, or Reflector Type Klystron

The Magnetron

24. Description
25. Electron Orbits
26. The Split-Anode Magnetron Operating in the Dynatron Mode (Habann Oscillator)
27. The Resonant Cavity Centimetric Magnetron
28. The Single Stream Steady State
29. Behaviour of the Space-Charge Cloud when the Anode Voltage Varies
30. The Electrometer Analogy
31. Modes of Oscillation
32. Initiation of Oscillation - The Energy Transfer Criterion
33. Maintenance of Oscillation - The Energy Supply Criterion
34. The Two Criteria Combined
35. Strapping
36. Effect of Anode Length, and of End-Plate Distance
37. Output Couplings
38. Cathode Features
39. Pulse Operation of Magnetrons
40. Magnetron Characteristics and Rieke Diagrams
41. Frequency Stability and Pulling Figure
42. Pre-Flumbing of Magnetrons

Control of Negative-Grid Oscillators

43. Stabilisation of Oscillators
44. Steady State Conditions
45. Effect of Supply Potentials on Steady State Conditions
46. Biasing Circuits
47. Self-Quenching Oscillators
48. The Super-Regenerative Receiver

Automatic Frequency Control

49. General
50. The Discriminator
51. Amplitude Discrimination
52. Phase Discrimination
53. Automatic Sweep Circuit
54. Practical AFC Circuits

CHAPTER 9

VALVES AS AMPLITUDE LIMITERS

1. General Principles
 2. Diode Limiting Circuits
 3. Limiting Circuits Employing Triodes or Pentodes
 4. Relative Merits of Triodes and Pentodes
- Applications of Amplitude Limiters
5. The production of Rectangular Pulses
 6. Elimination of Irregularities
 7. Pulse Selection
 8. Time-Constant Considerations Affecting the Output Circuit of a Limiting Amplifier

CHAPTER 10

RELAXATION OSCILLATORS AND RELAYS

1. Introduction
- Two-Valve Circuits
 2. The Multivibrator
 3. Cathode-Coupled Multivibrator
 4. Relay Circuits
- One-Valve Circuits
 5. Transitron Circuits
 6. Transitron Relaxation Oscillator
 7. Transitron Relay
 8. The Blocking Oscillator
- Circuits Employing the Miller Time-Base Principle
 9. General
 10. The Sanatron
 11. The Phantatron

CHAPTER 11

TIME BASES AND TIME-BASE GENERATORS

- Introduction
 1. General
 2. Linearity
 3. Flyback
 4. Time Bases Derived from a Source of Steady Voltage
- Time-Base Generators for Electric Deflection
 5. Fundamental Circuit of Capacitive Time-Base Generator
 6. The Charging Circuit of a Capacitive Time-Base Generator
 7. Methods of Improving Linearity of Charging
 8. Discharge Circuits
 9. Puckle Time-Base Generator
 10. Miller Time-Base Generator
 11. Time-Base Generators for Magnetic Deflection
12. Circular, Elliptical and Spiral Time Bases
13. Radial Time Base

CHAPTER 12

RECTIFYING AND CLAMPING CIRCUITS

1. Rectification
2. Simple Clamping Circuits
3. Switched Clamping Circuits

CHAPTER 13

THE PRODUCTION OF SHORT DURATION PULSES

1. Introduction
2. The Use of a Short Time-Constant C-R Circuit
3. The Use of a Ringing Circuit
4. The Use of a Pulse-Forming Network

CHAPTER 14

PULSE MODULATION OF TRANSMITTERS

1. Modulation Requirements
2. Modulation Methods
 - Grid Modulation
 3. Pulse Amplifier Method
 4. Cathode Follower Method
 5. Grid Self-Quenching
 6. Anode Self-Quenching
 7. Disadvantages of Grid Modulation
 - Anode Modulation
 8. General
 9. Hard Valve Modulator
 10. Hard Modulator Valves
 11. Anode Modulation Using a Delay Network in Conjunction with a Spark Gap or a Gas Triode
 - Pulse Transformers
 12. General
 13. Equivalent Circuit of a Pulse Transformer
 14. Pulse Transformer Connected Between a Modulator Valve and Magnetron
 15. Design Considerations
 16. Overswing Diode
 17. Pulse Transformer with Bifilar Secondary
 18. Use of Pulse Transformer in Conjunction with Hard Valve Modulators
 19. Use of Pulse Transformer in Conjunction with Delay Network
 - Hard Valve Modulator Circuits
 20. Pre-Modulator
 21. Boot-Strap Pre-Modulator
 22. Single Valve Pre-Modulator
 - Charging Delay Network from Steady Voltage Source
 23. Choke Charging
 24. Constant Current Charging
 25. Symmetrical Charging
 26. Typical Spark Gap Modulator
 - Charging Delay Network from an Alternating Voltage Source
 27. Advantages of Alternator Charging
 28. The Charging Circuit
 29. The Equivalent Charging Circuit
 30. Transients due to Line Discharge
 31. Graphical Derivation of Transient State
 32. Normal Transient Conditions
 33. Effect of Overswing
 34. Choice of the Ratio f_n/f_a
 35. Causes of Overswing
 36. Method of Triggering Discharge Valve
 - Voltage Multiplying Circuits
 37. General
 38. The Marx Circuit
 39. Blumlein Modulator

CHAPTER 15

NOISE

1. Introduction
2. Signal-to-Noise Ratio

3. Internal Noise
4. Thermal Noise
- Shot Noise
 5. General
 6. Diode Operating under Temperature-Limited Conditions
 7. Diode Operating under Space-Charge Limitation Conditions
 8. Negative-Grid Triode under Space-Charge Limitation Conditions
9. Partition Noise
10. Induced Grid Noise
11. Further Sources of Noise
- The Equivalent Noise Resistance of a Valve
 12. General
 13. Triode Valves
 14. Pentode Valves
15. Valve Noise under Feedback Conditions
16. Summary

CHAPTER 16

RECEPTION OF RADIO FREQUENCY PULSES

Introduction

1. General
2. Fourier Analysis
3. Distortion of RF Pulse During Amplification
4. The Effect of Noise
5. The Superheterodyne Principle
6. Typical 200 Mc/s Radar Receiver

Influence of Noise Factors on RF Amplification

7. General
8. The Ideal RF Amplifier
9. The Practical RF Amplifier : Noise Factor
10. Input Circuit Coupling with Noiseless First Valve
11. Input Circuit Coupling with Noisy First Valve
12. Effects of Transit Time and Cathode Lead Inductance
13. Operation at Higher Frequencies
14. The Grounded-Grid Triode

Frequency Conversion (Mixing)

15. General
16. Fundamental Principle of Frequency Conversion
17. The Single-Input Mixer
18. The Double-Input Mixer
19. The Two-Pole Converter
20. The Diode Mixer
21. The Effect of Diode Capacitance and Lead Inductance
22. The Crystal Mixer
23. Effect of Crystal Capacitance
24. Noise in Frequency Converters
25. Noise in Triode, Pentode and Hexode Mixers
26. Noise in Diode Mixers
27. Noise in Crystal Mixers
28. Local Oscillator Noise
29. The Overall Noise Factor of Crystal Mixer and IF Amplifier

Radar Receivers

30. General
31. The Overall Noise Factor of a Receiver
32. The Number of Signal Frequency Stages
33. Receiver for Metre and Decimetre Wavelengths
34. Receivers for Centimetre Wavelengths
35. The IF Amplifier

36. Manual Gain Control
37. A.G.C.
38. Paralysis
39. Receiver Suppression
40. Temporal Gain (Swept Gain)

CHAPTER 17

AERIALS

1. Fundamental Considerations

Transmitting Aerials

2. Type of Wave Emitted by a Transmitting Aerial
3. Field-Strength Diagram of an Aerial
4. Field-Strength Diagram of Half-Wave Aerial
5. Isotropic Radiator
6. Energy Density in an Electromagnetic Wave
7. Power Relations for a Plane Electromagnetic Wave
8. Power and Field Relations for an Isotropic Radiator
9. Power and Field Relations for any Aerial
10. Power Gain of Half-Wave Aerial
11. Input Impedance
12. Q-Factor of a Resonant-Length Aerial
13. Factors Affecting the Resonant Length

Receiving Aerials

14. General
15. Power Relations for an Isotropic Receiving Aerial
16. Power Relations for any Receiving Aerial

17. Transmission and Reception

Propagation and Reception of Short Waves

18. Scattering by an Obstacle
19. Range of a Radar Set
20. Reflection from Ground and Sea
21. Horizontally Polarised Waves Incident on Ground
22. Horizontally Polarised Waves Incident on Sea
23. Reflection of Vertically Polarised Waves
24. Effect of Flat Earth on Field Strength Diagram (Horizontal Polarisation)
25. Effect of Flat Earth on Field Strength Diagram (Vertical Polarisation)
26. Summary of Effect of Flat Surface
27. Gap Filling
28. Effect of Earth's Curvature : Refraction

Common Types of Aerial Arrays

29. Broadside Array
30. Linear Broadside Array
31. Complete Broadside Array
32. Use of Reflecting Screen Behind Broadside Array
33. Wire Netting Reflector Screens
34. Power Gain of Broadside Array
35. Feeding Arrangements for Broadside Arrays
36. Example of a Complete Broadside Array
37. Tapered Feed to Linear Broadside Arrays
38. Beam Swinging with Linear Broadside Arrays
39. End-Fire Arrays
40. Comparison of End-Fire and Broadside Arrays
41. Parasitic Director
42. End-Fire Array with Parasitic Directors (Yagi)
43. Parasitic Reflector
44. Use of Folded Half-Wave Aerial
45. Bandwidth of Arrays

46. Effect of Ground on Aerial Arrays
47. Gap-Filling by means of Phasing
- Microwave Aerials
 48. Paraboloidal Mirrors
 49. Primary Feeds for Paraboloidal Mirrors
 50. Beam Swinging with Paraboloidal Mirrors
 51. Specially Shaped Mirrors
 52. Cheese Type of Mirror
 53. Band-Width of Mirror-Type Aerials
 54. Slots in Waveguides
 55. Slotted Linear Arrays
 56. Slotted Waveguide with Mirror
 57. Microwave Beacon Aerials
 58. Effect of Ground on Microwave Aerials
59. Conclusion

CHAPTER 18

CONTROL SYSTEMS

1. Introduction
- Data Transmission Systems (Non-Resetting)
 2. General
 3. Performance of Transmission Systems
 4. Step-by-Step Systems for Transmitting Position Data
 5. Potentiometer and Voltmeter Systems
 6. Ring-Potentiometer Systems
 7. Single-Phase AC Systems
 8. Selsyn Systems
 9. Construction of Selsyn Transmitter
 10. Selsyn Receivers
 11. The Magslip Receiver
 12. Coincidence Indicating Systems
- Servo Systems
 13. General
 14. Computing Devices (Converting and Calculating Elements)
 15. Characteristics of the load
 16. Servo Motors
 - DC Motors
 - Two-Phase Induction Motor
 - Ward-Leonard System
 - Metadyne System
 - Speed-Control System
 17. Performance of Servo Systems
 18. Definitions and Assumptions (Applying to a Displacement-Displacement Servo)
 19. Simple Error Control
 20. Limitations of Simple Error Control
 21. The Use of Derivative and Integrated Error Control
 22. Speed Control (First Order Servo)
 23. Second Order Servo with Zero Velocity Lag

CHAPTER 19

POWER SUPPLIES

1. Introduction
- Rectifier Circuits
 2. General
 3. Single-Phase Circuits

- 4. Polyphase Circuits
 - Filter Circuits
 - 5. General
 - 6. Series-Choke Input Filter
 - 7. Shunt-Condenser Input Filter
 - 8. C-R Filter
 - Stabilising or Regulating Circuits
 - 9. Requirements
 - 10. Neon Valve Stabilisers
 - 11. Thermistor Stabilisers
 - 12. Hard-Valve Stabilisers
 - Regulated Transformers and AC Stabilisation
 - 13. General
 - 14. Alternating Current Regulator
 - 15. Alternating Voltage Regulator : Motor Circuit
 - 16. Alternating Voltage Regulator : Saturable Reactor Circuit
 - 17. Automatic Potentiometer (Electronic Potentiometer) Earthing circuit
-

CHAPTER 1

LINEAR CIRCUIT ANALYSIS

1. INTRODUCTION

This chapter summarises those items of conventional electric circuit theory which form the basis of the analysis in the remainder of Part II. It consists largely of statements of results, and other texts should be consulted if proofs are required. For present purposes the circuit is regarded as a collection of components (or circuit elements) interconnected by a mesh of short conductors and stimulated electrically at localised places, the whole assembly being so compact that electro-magnetic effects can be assumed to be propagated instantaneously within this self-contained system. The behaviour of the circuit is described primarily in terms of the potential differences existing between pairs of "points" in the circuit and of the currents which flow through "points" in the circuit; the word "point" here usually signifies the cross-section of a conductor.

The symbol for a potential difference (or voltage) is v , and that for a current is i .

The subject matter of this chapter is dealt with in the following standard Service Reference Books :-

Admiralty Handbook of Wireless Telegraphy Vol. 1.

(BR 229)

Signal Training (26/Manuals/1451 and 1577) Vol. II Parts I and II.

Royal Air Force Signal Manual Part II.

(AP 1093)

References to particular passages are given in the appropriate sections of this chapter.

In particular the matter of Sec. 12 should be taken as supplementing, not replacing, the description given in the standard Service Reference Books. The appropriate references are :-

BR 229. Chap. V. (Paras. 272-348).
26/Manuals/1451. Chap. XXV. (Secs. 231-240).
26/Manuals/1577. Chaps. II and III. (Secs. 9-17).
AP 1093. Chap. V.

2. CIRCUIT LAWS AND CONVENTIONS

Circuit theory is founded on two laws attributed to Kirchhoff. One governs the relations between currents and the other the relations between potential differences in a circuit. These laws may be stated as follows :-

- (i) The algebraic sum of the currents entering a point is zero at any instant.
- (ii) The algebraic sum of the potential differences encountered in traversing a closed loop is zero at any instant.

Thus in Fig. 1, application of the first law with regard to the currents $i_1, i_2 \dots i_7$ provides the following independent relations :-

$$(i) \begin{cases} i_1 - i_2 - i_4 + i_7 = 0 \\ i_2 + i_3 - i_5 - i_7 = 0 \\ i_1 + i_3 - i_6 = 0 \end{cases}$$

Similarly the second law, applied in respect of the potential differences $v_1, v_2 \dots v_{10}$, provides the following independent relations :-

$$(ii) \begin{cases} v_1 - v_4 - v_7 - v_9 = 0 \\ v_2 - v_5 - v_{10} = 0 \\ v_3 - v_6 - v_8 - v_9 = 0 \\ v_7 - v_8 + v_{10} = 0 \end{cases}$$

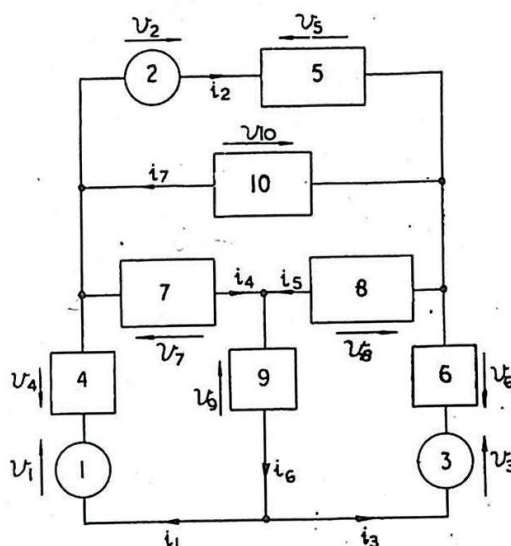


Fig. 1 - Typical circuit containing active and passive elements.

Although other similar relations may be available they will be found to be implicit in those stated.

In the diagram three types of circuit elements are indicated:

- a thin line denoting an ideal connector between other elements, and providing no impediment to the current;
- a rectangular box denoting a passive element, i.e. one which normally impedes the current, and
- a circular box denoting an active element, i.e. one which normally drives the current.

The directions of reference for currents and potential differences are indicated by arrows; those for the potential differences are alongside the elements to which they refer, the tip of an arrow being taken to indicate the terminal of higher potential, and those for the currents are superimposed on the connectors, their direction indicating the flow. (This is the opposite direction to that of the movement of the electrons). It is emphasised that these directions are solely for reference and correspond with actual voltages and currents only when these emerge from the analysis with positive values. It may be noticed that in a simple circuit having only one generator the voltage arrow appears to assist the current arrow through the active element, and to oppose it through passive ones, although this useful relation is only a consequence of the above conventions.

Any closed path in a network is known as a Mesh, and the application of Kirchhoff's laws can be simplified by using the conception of Mesh Current Components to replace currents through the elements. Fig. 2 represents the same circuit as Fig. 1 when redrawn from this point of view; (the voltage arrows are omitted for the sake of clarity). The current in any direction through a circuit element is the algebraic sum of the components in that direction contributed by those meshes of which it forms a part. Thus, by a comparison of Figs. 1 and 2 the following relations emerge :-

$$\begin{aligned}
 i_1 &= i_A \\
 i_2 &= i_B \\
 i_3 &= i_C \\
 i_4 &= i_A + i_D \\
 i_5 &= i_C - i_D \\
 i_6 &= i_A + i_C \\
 i_7 &= i_B + i_D
 \end{aligned}$$

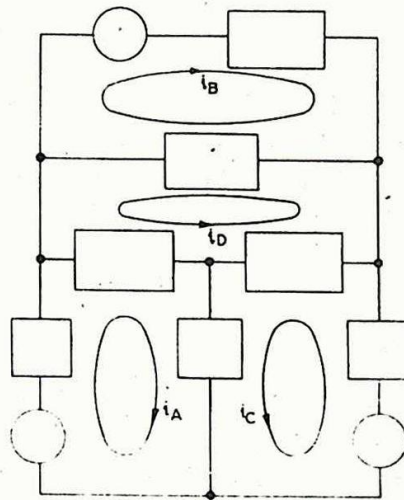


Fig. 2 - Circuit showing mesh currents.

These will be found to satisfy equations (i), and signify that by the introduction of mesh current components Kirchhoff's first law is automatically satisfied and the number of variables reduced. The components also simplify network analysis for other reasons; but it should be remembered that they exist only as a convenience in analysis, and cannot be endowed with a separate physical existence in circuits.

Substitution Theorem

This useful theorem, of general application, is self-evident, and may be stated as follows :-

A part of any network, coupled only to the remainder through its terminals, may be replaced without affecting the behaviour of the remainder of the network by a system of ideal generators of voltage (or current) which are arranged so as to preserve the previously existing potential differences (or currents) at the terminals. (The conception of an ideal generator of voltage (or current) is considered in Sec. 4.)

References :

BR 229. Paras. 72 and 73.
26/Manuals/1451 Sec. 136.
AP 1093. Chap. 1. Paras. 31-38.

CHARACTERISTICS OF CIRCUIT ELEMENTS

3. Passive Elements

Circuit problems cannot be solved simply from a knowledge of circuit laws; it is also necessary to know the characteristic properties of each element in the circuit so as to be able to predict its behaviour when subjected to any specific variation of voltage or current. Although such properties may on occasion be very complicated, and expressible only in the form of experimental data, tabulated or presented by means of graphs, there is a wide range of elements for which much simpler specification is available. Thus it is found that for one class of circuit elements the current i is always proportional to the applied voltage v . Where this obtains the elements are described as resistive (or purely resistive) and are known as resistors. The ratio $\frac{v}{i}$

called the resistance of the element, denoted by R . The appropriate

circuit symbol is shown in Fig. 3(a). The result

$$\frac{v}{i} = \text{constant (i.e. } R \text{ is constant)}$$

is known as Ohm's law.

For another type of element the current i is always proportional to the rate of change of the applied voltage, $\frac{dv}{dt}$. Such

elements are said to be capacitive and are known as condensers (or capacitors). The ratio $i/\frac{dv}{dt}$ is called the capacitance of

the element, denoted by C . The appropriate circuit symbol is shown in Fig. 3(b).

The result :-

$$\frac{i}{\frac{dv}{dt}} = \text{constant}$$

(i.e. C is constant)

has no particular name; it is the law for a condenser, and is usually quoted in the form

$$\frac{q}{v} = C$$

where q is the charge on one plate.

For yet another type of element the applied voltage is always proportional to the rate of change of current, $\frac{di}{dt}$. Such elements are

said to be self-inductive, and are known as coils (or self-inductors). The ratio $v/\frac{di}{dt}$ is called the self-inductance (or simply the inductance)

of the element, denoted by L . The appropriate symbol is shown in Fig. 3(c).

So far all elements considered have had one pair of terminals; but there is another type of inductive element with two pairs, having the property that if the current through one pair (primary) varies, then a voltage is established between the other pair (secondary), which, on open circuit, is always proportional to the rate of change of the primary current. Such elements are known as transformers (mutual inductors) and the ratio $\frac{v(\text{secondary})}{\frac{di}{dt}(\text{primary})}$ is called the mutual inductance, denoted by M . The

appropriate symbol is shown in Fig. 3(d).

These last two results, of the form

$$\frac{v}{\frac{di}{dt}} = \text{constant (i.e., } L \text{ or } M \text{ is constant)}$$

are a consequence of Faraday's Law of Electromagnetic Induction, which states that the induced EMF is proportional to the rate of change of the flux turns threading a circuit.

Actually no elements conform exactly and under all circumstances to the ideal concepts of resistance, capacitance and inductance

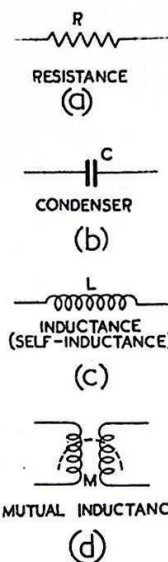


Fig. 3 - Circuit symbols

which have been defined, though some may do so nearly enough for practical purposes within a more or less restricted range of circumstances. Other elements may exhibit behaviour which can be simulated by combinations of ideal elements, the degree of simulation increasing with the complexity of the "equivalent network".

Thus in Figs. 4(a) and (b) are shown three successive approximations to simulate the properties of a coil of wire and of a condenser, respectively.

It should be clearly understood that such representations serve a most useful purpose as equivalents, but that identification of parts of the representation with parts of the component represented is by no means always permissible, and that such points in the equivalent circuit as the junctions of R and L in Fig. 4(a) have no physical existence.

Circuit elements which may be represented adequately by an equivalent network made up of unvarying resistance, inductance and capacitance are known as Linear Elements, and circuits whose passive elements are all linear are known as linear circuits or networks. Subject to limitations, the theory of linear circuits may also be applied to circuits in which the fundamental parameters, resistance, inductance and capacitance, are functions of applied frequency or of time: specification of these limitations is beyond the scope of the text, except to state that the more gentle the variation of parameters, as compared with the mode of exciting the circuit, the more accurate is the analysis.

References

BR 229. Paras. 58-68 and 128-183.

26/Manuals/1451. Secs. 127-133, 185-202 and 60-72.

AP 1093. Chap. 1. Paras. 26-30 and 47-58.

Chap. 2. Paras. 24-43.

4. Active Elements

Active circuit elements are primarily sources of electromotive-force and are ultimately responsible for all the electrical behaviour displayed by actual circuits. The voltage developed at the terminals of an active element is not the same in general as the EMF generated, the difference between these quantities being ascribed to losses in the element itself and arising from internal impediments to the current established.

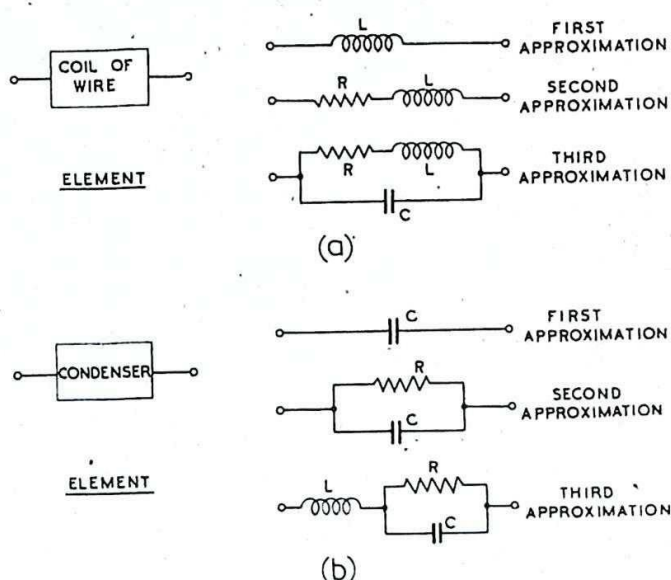


Fig. 4 - Equivalent networks : successive approximations.

For a comprehensive circuit analysis it is therefore necessary to know the voltage which would be developed at the terminals of each particular generator under all

conditions of load.

Just as it has been found convenient to describe the properties of a passive element in terms of an equivalent network, so it is helpful to replace an active element by an equivalent combination of elements

in which the primary action (generation of EMF) is ascribed to an ideal active element, devoid of loss, associated with a passive element, representing the internal impedance; (see Fig. 5).

For many practical active elements the internal impedance may be represented by a linear network, and further, the characteristics of the ideal generator may be found to be independent of load; both of these qualifications materially simplify circuit analysis. The ideal active element of Fig. 5 may be termed a Constant Voltage Generator, as it delivers an EMF as stated under all conditions of load. An alternative equivalent representation is shown in Fig. 6, in which the ideal active element may be described as a Constant Current Generator, in that its function is to maintain a current as stated under all conditions of load.

The idea of a constant current generator may seem rather curious to those who naturally regard potential difference as cause and current as effect. In circuit analysis, however, it is only the relations between these two quantities which have special significance, and it is not necessary to assert that either notion is more fundamental than the other. Thus the notions of ideal voltage generators and ideal current generators are equally feasible.

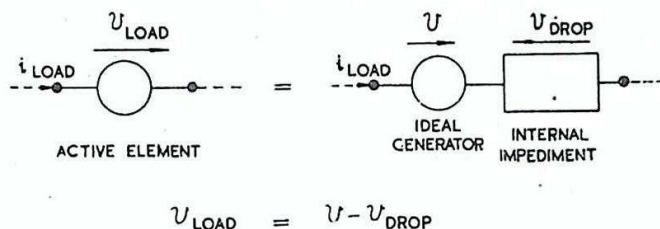


Fig. 5 - Representation of an inactive element : constant voltage equivalent network.

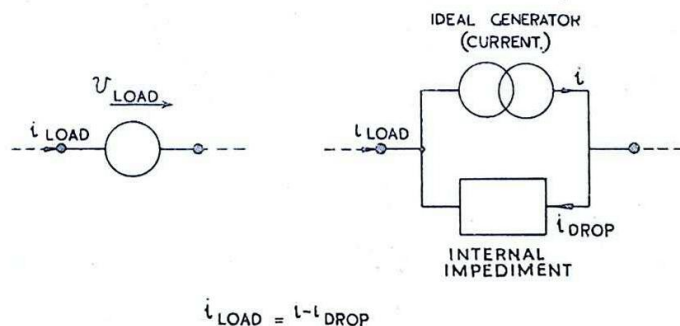


Fig. 6 - Representation of an active element: constant current equivalent network.

References

BR 229. Paras. 49-54.

26/Manuals/1451. Sec. 117.

AP 1093. Chap. 1. Paras. 17-22.

LINEAR CIRCUITS

5. General

Linear circuits are those in which all passive elements (including those representing the internal impediments of active elements) can be represented by equivalent combinations of resistance, inductance and capacitance. Circuits to which this restriction cannot be applied are generally termed Non-Linear (or Non-Ohmic), and the analysis of their behaviour is usually complicated.

Certain results hold in the field of linear circuits which do not in general permit of wider application. These are stated below in the form of circuit theorems.

6. Superposition Theorem

In a linear network the current at any point (or the potential difference between any two points) may be regarded as the algebraic sum of the corresponding currents (or potential differences) produced as each generator in the network is considered to operate independently with all the others suppressed.

To suppress a generator, in this sense, means to reduce its output to zero without affecting its internal impediment; this may be regarded as equivalent to short-circuiting the ideal voltage generator (or open-circuiting the ideal current generator) in the equivalent circuit of an active element; (see Figs. 5 and 6).

7. Reciprocity Theorem

In any passive linear network the current produced at one point as a result of inserting an ideal voltage generator at another is identical with the current which would be produced at the second point by inserting the same generator at the first. Similarly, the potential difference arising between any pair of points from bridging any other pair with an ideal current generator is identical with the potential difference which would arise between the second pair of points by bridging the first pair with the same generator.

REPRESENTATION OF SINUSOIDAL CURRENTS

AND VOLTAGES IN LINEAR CIRCUITS

8. General

Throughout this volume a capital letter (e.g. V) is used to denote any voltage or current which in the circumstances may be taken as remaining constant over the interval of time considered. Where the quantity under discussion varies it is denoted by a small letter, (e.g. v). Thus,

$$v = f(t)$$

is the relation indicating that the value of v at any instant t can be

expressed as a function of t . In the particular case of a sinusoidal function, v might satisfy the relation

$$v = \hat{v} \sin \omega t,$$

\hat{v} denoting the amplitude, or peak value, of v , and ω the angular frequency, which is $2\pi f$, where f is the frequency of the oscillatory voltage.

For the remainder of this chapter we are concerned with the analysis of linear networks which are subjected to steady sinusoidal voltages, emphasising in particular those points which are useful in subsequent chapters and which are not stressed in the references quoted. The behaviour of circuits which are not linear or which are subjected to other kinds of voltage-variations is discussed in other chapters.

The importance of the sinusoidal voltage in this field is due to the ease with which the results for all types of circuit elements can be specified in simple form. Thus, in any linear network subjected to one or more sinusoidal voltages of the same frequency the current at any point and the voltage between any two points are also sinusoidal (of the same frequency) and can be specified in terms of amplitude and phase, measured with respect to one of the generators. This form of specification also avoids an undue emphasis on the instantaneous values, which have rarely any special interest, notwithstanding the fact that they represent the actual events occurring.

In building up a sinusoidal analysis of circuit behaviour, the trigonometry associated with instantaneous values is commonly avoided by the use of vector representation, in which instantaneous values are regarded as the instantaneous projections of rotating vectors. Such vectors are used to represent voltages and currents, length representing amplitude and direction representing phase. Phase difference is represented exactly by difference in direction of two such vectors.

By the use of complex numbers the geometry of such vector diagrams can be expressed in algebraic form. For instance,

$$i = \hat{i} \mathcal{E} j(\omega t + \phi)$$

represents a vector diagram in which a vector of length \hat{i} rotates at angular velocity (or frequency) ω , passing through the direction ϕ at zero time: Now

$$\hat{i} \mathcal{E} j(\omega t + \phi) = \hat{i} \left[\cos(\omega t + \phi) + j \sin(\omega t + \phi) \right],$$

the term $\hat{i} \cos(\omega t + \phi)$, (known as the Real Part) denoting the projection on the horizontal axis, and the term $j \hat{i} \sin(\omega t + \phi)$ (known as the Imaginary Part), the projection on the vertical axis. This enables vector treatment to be translated into the actual instantaneous values at any stage by considering the real parts of the expressions for voltage and current. The use of complex numbers in expressions infers that voltage and current are also specified in similar terms, and the rewording of the results in terms of real voltages and currents is often left to the reader; this is a common practice in electrical text books.

For convenience the representation on a diagram of such a vector as

$$v = \hat{v} \mathcal{E} j\omega t$$

is denoted by an arrow, thus; \vec{v} .

References

BR 229. Paras. 339-348

26/Manuals/1577, Chap. II. Appendix I.

AP 1093. Chap. V.

9. Complex Quantities

Fig. 7 represents a vector \vec{r} which is shown in relation to a pair of perpendicular axes; the horizontal axis is called the Real axis and the vertical axis the Imaginary axis. Vectors drawn parallel to the imaginary axis are prefixed with j to denote this direction. Thus \vec{a} represents a step of amount a parallel to the real axis and \vec{jb} a step of amount b parallel to the imaginary axis. Combining these steps the resultant becomes $\vec{a} + \vec{jb}$ represented in the diagram by \vec{r} . Hence we can write

$$\vec{r} = \vec{a} + \vec{jb} \quad (\text{Cartesian form}).$$

In analysis the arrows are omitted, so that the expression is written

$$r = a + jb$$

Alternatively, \vec{r} is a vector of length $|r|$ making a direction θ with the real axis, and can be thought of as a vector of length $|r|$, originally lying along the real axis, which has been rotated through an angle θ . Two notations can be used to denote this process:

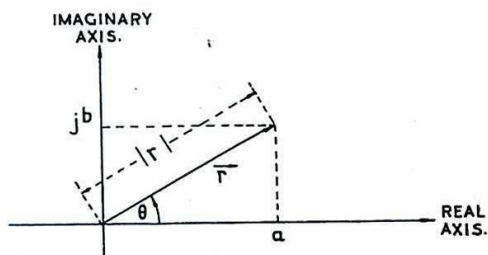


Fig. 7 - Representation of a vector.

$$(i) \quad r = |r| \angle \theta \quad (\text{polar form})$$

$$\text{or} \quad (ii) \quad r = |r| \angle j\theta \quad (\text{exponential form})$$

The latter form is used in this volume.

We can thus write

$$\begin{aligned} r = a + jb &= |r| \angle j\theta \\ &= |r| \cos \theta + j |r| \sin \theta; \end{aligned}$$

$$\text{whence,} \quad a = |r| \cos \theta$$

$$\text{and} \quad b = |r| \sin \theta$$

$$\text{Conversely, } |r| = \sqrt{a^2 + b^2}$$

$$\text{and } \tan \theta = \frac{b}{a}.$$

We may use complex numbers to represent the effects produced by sinusoidal voltages and currents if the numbers are carefully interpreted as explained in Sec. 8. For these applications, r becomes v or i , whilst $|r|$ becomes $|v|$ or $|i|$. Since $|v|$ or $|i|$ is equal to the maximum or peak value of the real part of the voltage or current, it is more convenient to use the symbols \hat{v} and \hat{i} in place of $|v|$ and $|i|$.

For both current and voltage, θ varies uniformly, executing one rotation per cycle, and has angular velocity $\omega = 2\pi f$. Thus,

$$v = \hat{v} \mathcal{E} j(\omega t + \phi_1) \quad \text{and} \quad i = \hat{i} \mathcal{E} j(\omega t + \phi_2)$$

represent a voltage and a current of amplitudes \hat{v} and \hat{i} and phases $\omega t + \phi_1$ and $\omega t + \phi_2$ respectively, the voltage leading the current in phase by the angle

$$\begin{aligned} & \omega t + \phi_1 - (\omega t + \phi_2) \\ = & \phi_1 - \phi_2. \end{aligned}$$

We may also use complex numbers to denote the ratio between the vectors representing voltage and current. Thus

$$\begin{aligned} \frac{v}{i} &= \frac{\hat{v} \mathcal{E} j(\omega t + \phi_1)}{\hat{i} \mathcal{E} j(\omega t + \phi_2)} \\ &= \frac{\hat{v}}{\hat{i}} \mathcal{E} j(\phi_1 - \phi_2). \end{aligned}$$

This new vector $\frac{v}{i}$ is stationary and is usually denoted by $z = Z \mathcal{E} j\phi$

$$\text{where } \frac{\hat{v}}{\hat{i}} = Z$$

$$\text{and } \phi = \phi_1 - \phi_2.$$

z can also be written in the cartesian form

$$z = R + jX,$$

where R and X correspond to the Resistance and Reactance encountered in the usual analysis.

Care must be taken in using complex numbers to represent voltages and currents in calculations involving power. For example, if the instantaneous value of the voltage applied to a load is indicated by the real part of

$$v = \hat{v} \mathcal{E} j(\omega t + \phi_1)$$

and that of the load current by the real part of

$$i = \hat{i} \mathcal{E} j(\omega t + \phi_2)$$

then the instantaneous power consumed by the load is given not by the real part of the product vi but by the product of the real parts of v and i considered individually. Its mean value is $\frac{1}{2} \hat{v} \hat{i} \cos(\phi_1 - \phi_2)$. $\cos(\phi_1 - \phi_2)$ is called the Power Factor of the load. (In general the Power Factor is defined as the ratio of the mean power consumed by the load to the product of the RMS (Root Mean Square) amplitudes of the load voltage and current. For sinusoidal variations the RMS value (V or I) is $\frac{1}{\sqrt{2}}$ times the maximum value, so that

$$VI = \frac{\hat{v}}{\sqrt{2}} \cdot \frac{\hat{i}}{\sqrt{2}} = \frac{1}{2} \hat{v} \hat{i} \cos(\phi_1 - \phi_2)$$

10. Vector Diagrams

Fig. 8 represents a voltage given by

$$v = \hat{V} \epsilon j\omega t$$

as a vector of length \hat{V} and rotating at an angular velocity ω . When several vectors are to be compared, representing sinusoidal voltages or currents of the same frequency, it is usual to replace the diagram of rotating vectors by referring all the vectors to one of the set as a reference vector. This is normally drawn along the horizontal axis, and each of the other vectors is given the appropriate length and phase compared with those of the reference vector; (Fig. 9). All vector diagrams in this volume representing voltages or currents are of this relative type unless otherwise stated.

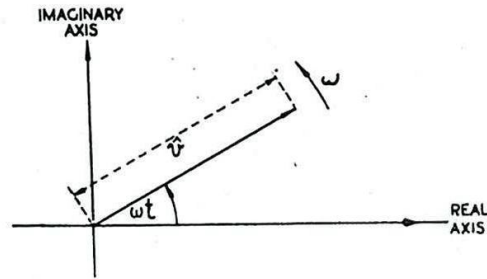


Fig. 8 - Rotating vector.

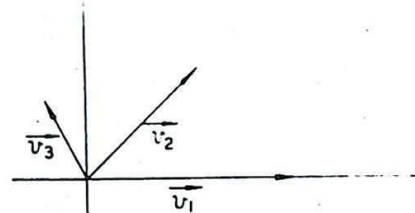


Fig. 9 - Relative vector diagram.

For convenience, when comparing vectors a vector may be transplanted to any origin in its plane, since the only significant quantities are length and direction, and provided these remain the same the significance is unaltered.

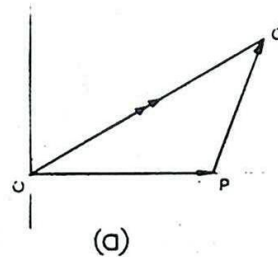
Vectors are added as indicated in Fig. 10(a);

$$\vec{OP} + \vec{PQ} = \vec{OQ}.$$

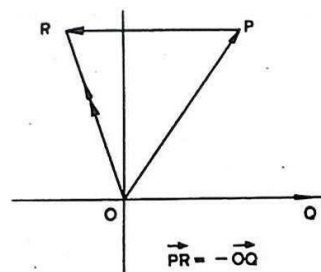
To subtract one vector from another the vector to be subtracted is reversed in direction and the reversed vector is then added to the other.

E.g., referring to Fig. 10(b),

$$\begin{aligned} \vec{OP} - \vec{OQ} &= \vec{OP} + \vec{PR} \\ &= \vec{OR}. \end{aligned}$$



(a)



(b)

Fig. 10 - Addition and subtraction of vectors.

IMPEDANCES AND ADMITTANCES

11. General

As explained in Sec. 8, if v and i are complex quantities representing respectively the sinusoidal voltage applied to a linear two-terminal network and the current which flows through it, the ratio $\frac{v}{i}$ is a complex number called the complex impedance, or, more simply, the impedance of the network. It is denoted by the symbol z . The phase angle of this impedance, which is the angle by which the voltage vector leads the current vector, is denoted by ϕ , and the magnitude of z by Z .

Thus

$$z = Z \angle \phi.$$

(In this volume the phrase "magnitude of the impedance" is used to denote Z ; the common but somewhat confusing practice of using the word impedance to denote both the complex number z and its magnitude Z is thereby avoided). For a passive network ϕ may have any value between $\pm 90^\circ$ ($\pm \frac{\pi}{2}$ radians). If $0 < \phi \leq 90^\circ$ the network is inductive; for $-90^\circ \leq \phi < 0$ the network is capacitive.

If $\phi = \pm 90^\circ$ the impedance is purely reactive and may be denoted by $z = jX$, where X is the reactance. Such an impedance is presented by an ideal coil of inductance L , where $X = \omega L$, ($\phi = +90^\circ$), or by an ideal condenser of capacitance C , where $X = -\frac{1}{\omega C}$, ($\phi = -90^\circ$).

In practice coils and condensers possess resistance, and the impedance is correspondingly modified; (see Sec. 3).

In general the series resistance R is the real part and the series reactance X the imaginary part of z , so that $z = R + jX$.

The steady state of a two-terminal network in response to a sinusoidal voltage of given frequency is determined exactly when the magnitude Z and phase angle ϕ of the impedance are known.

Alternatively, it may be simpler, particularly when dealing with parallel combinations of networks, to conduct the analysis in terms of admittances rather than impedances. The ratio i/v is a complex number, called the (complex) admittance of the network, and is denoted by y . Hence $y = 1/z$ and the magnitude of the admittance is given by $Y = 1/Z$, so that $y = Y \angle -\phi$.

The real part of y is called the Conductance and is denoted by G , and the imaginary part is called the Susceptance, denoted by B . Thus

$$y = G + jB.$$

The susceptance of an ideal coil of inductance L is $-1/\omega L$, and that of an ideal condenser of capacitance C is ωC . For an ideal resistor of resistance R , $G = 1/R$.

Further comparisons between admittances and impedances are made in Sec. 15.

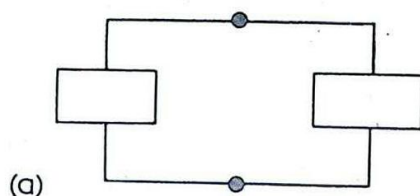
It should be borne in mind that knowledge of the magnitude and phase angle of an impedance (or an admittance) gives no knowledge of the nature of the components forming a two-terminal network. It merely indicates the overall effect at the frequency considered.

12. Helmholtz's, Thevenin's or Norton's Theorem

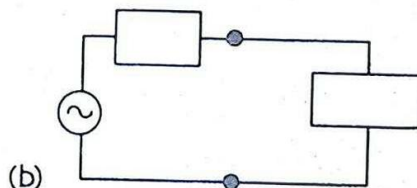
This theorem, variously attributed, wholly or in parts, to the above-named scientists, may be stated as follows :-

If a linear network can be divided into two parts, coupled only by a pair of conductors, then either part (removed network) may be replaced, without affecting the behaviour of the other part (remaining network), by the substitution of

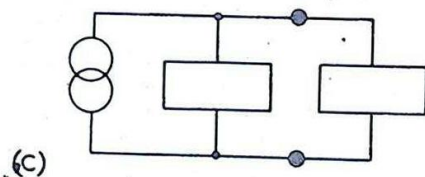
- (i) A constant voltage generator, giving the open-circuit voltage at the terminals of the removed part, in series with its output impedance; (Fig. 11 (b)); or
- (ii) A constant current generator, giving the short-circuit current at the terminals of the removed part, shunted by its output impedance; (Fig. 11 (c)).



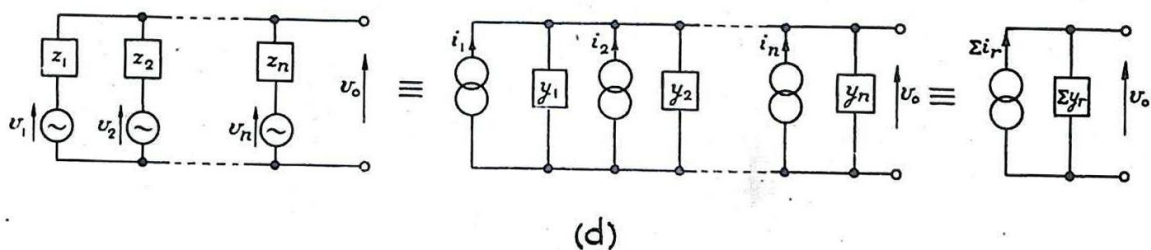
(a)



(b)



(c)



(d)

Fig. 11 - Helmholtz's, Thevenin's, or Norton's Theorem; Schematic representation.

The output impedance is defined as the impedance looking into the terminals of the removed network when all its generators are suppressed; (see Sec. 6 for this use of the term "suppress").

This theorem can be extended to networks in which the two parts referred to are coupled by any number of conductors.

It is legitimate to alter the make-up of the remaining network after substitution has been made for the removed network, as the latter is replaced by an equivalent combination of elements which are independent of the former. (Note that such variations cannot be made when the Substitution Theorem (Sec. 2) is employed, without respecifying the substitute generators).

As an example of the above theorem consider the circuit shown in the left-hand diagram of Fig. 11 (d). Each shunt branch may be considered as a constant-voltage generator (v_r) in series with its output impedance z_r . To calculate the output voltage v_o replace each branch by the equivalent representation i.e. a constant-current generator, giving the short-circuit current i_r , shunted by the output admittance y_r . The resultant current is $\sum i_r$ and the resultant admittance $\sum y_r$.

$$\text{Hence } v_o = \frac{\sum i_r}{\sum y_r}.$$

$$\text{But } i_r = v_r / z_r$$

$$\text{and } y_r = 1 / z_r.$$

$$\text{Hence } v_o = \frac{\sum \frac{v_r}{z_r}}{\sum \frac{1}{z_r}}$$

The above result holds equally for DC, and may then be quoted in the form

$$v_o = \frac{\sum V_r / R_r}{\sum 1 / R_r}$$

13. Star-Delta Transformation

This useful transformation is illustrated in Fig. 12. In the steady state any three-terminal linear network may be replaced by either of the forms shown at (a) and (b); the arrangement (a) is called a Star (Y) network, and that at (b) a Delta (Δ) network. If at any frequency the impedances forming the Star-network are z_1 , z_2 and z_3 , and those forming the Delta-network are z_a , z_b , and z_c , as shown, the conditions for the networks to be equivalent may be written in either of the two following forms :-

$$z_a = \frac{z_1 z_2 + z_2 z_3 + z_3 z_1}{z_1}$$

$$z_b = \frac{z_1 z_2 + z_2 z_3 + z_3 z_1}{z_2}$$

$$z_c = \frac{z_1 z_2 + z_2 z_3 + z_3 z_1}{z_3}$$

or

$$z_1 = \frac{z_b z_c}{z_a + z_b + z_c}$$

$$z_2 = \frac{z_a z_c}{z_a + z_b + z_c}$$

$$z_3 = \frac{z_a z_b}{z_a + z_b + z_c}$$

In filter theory similar networks are often denoted by the terms T (for Star) and π (for Delta).

14. Series Networks

It may be convenient to represent a two-terminal network by a simple series arrangement having the same impedance as the actual circuit. If this simple circuit consists of a resistance R_s and a reactance X_s in series, as shown in Fig. 13, we have

$$z = R_s + jX_s$$

$$Z^2 = R_s^2 + X_s^2$$

$$\text{and } \tan \phi = \frac{X_s}{R_s}$$

$$\text{Since } \cos \phi = \frac{R_s}{Z},$$

$$Z = R_s \sec \phi.$$

The power factor F of the network is equal to $\cos \phi$.

For an inductive network, represented as a resistor of resistance R_s in series with an ideal coil of inductance L_s ,

$$X_s = \omega L_s, \text{ so that}$$

$$\tan \phi = \frac{\omega L_s}{R_s}$$

For a capacitive network represented as a resistor of resistance R_s in series with an ideal condenser of capacitance C_s ,

$$X_s = -\frac{1}{\omega C_s}, \text{ so that}$$

$$\tan \phi = -\frac{1}{\omega C_s R_s}$$

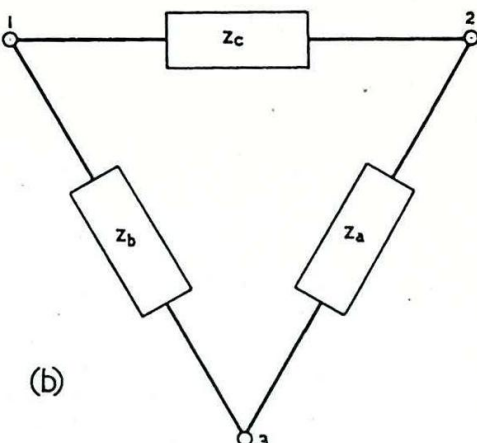
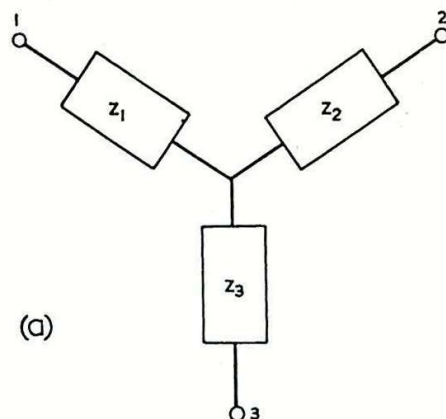


Fig. 12 - Star-delta transformation.

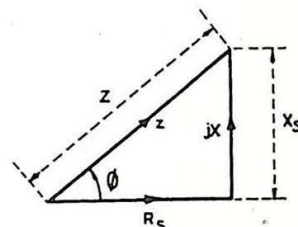
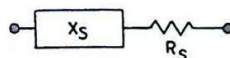


Fig. 13 - Simple series network.

15. Parallel Networks

It may be convenient to represent a network by a simple parallel arrangement having the same impedance as the actual circuit. If this simple circuit consists of a conductance G_p and a susceptance B_p in parallel, as shown in Fig. 14, we have

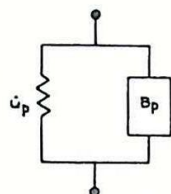
$$y = G_p + jB_p$$

$$Y^2 = G_p^2 + B_p^2$$

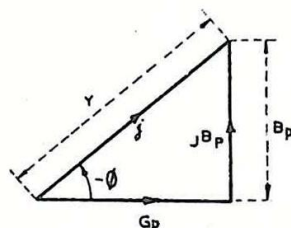
$$\text{and } \tan \phi = -\frac{B_p}{G_p}.$$

$$\text{Since } \cos \phi = \frac{G_p}{Y},$$

$$Y = G_p \sec \phi.$$



(a)



(b)

Fig. 14 - Simple parallel network.

The power factor is, as before, $\cos \phi$.

For an inductive network, represented as a resistor of reactance R_p in parallel with an ideal coil of inductance L_p ,

$$G_p = \frac{1}{R_p} \text{ and } B_p = -\frac{1}{\omega L_p}, \text{ so that}$$

$$\tan \phi = \frac{R_p}{\omega L_p}.$$

For a capacitive network represented as a resistor of resistance R_p in parallel with an ideal condenser of capacitance C_p ,

$$B_p = \omega C_p, \text{ so that}$$

$$\tan \phi = -\omega C_p R_p.$$

16. Q - Factor of a Component

For a single component the ratio $\frac{|\text{Reactance}|}{\text{Resistance}}$ is called the

Q - factor of the component. A practical coil can be represented approximately by an ideal inductance L in series with the resistance R of the coil; (see Fig. 4(a)). In this case the Q of the coil is $\frac{X_L}{R} = \frac{\omega L}{R}$. It follows from Secs. 14 and 15 that $Q = |\tan \phi|$;

also that if the component were to be represented for convenience as a pure susceptance in parallel with a pure conductance, Q would be given by $\frac{|\text{Susceptance}|}{\text{Conductance}}$. Since for a condenser this latter representation is more common, as indicated in Fig. 4(b), the Q - factor of

such a condenser is given by

$$Q = \frac{\omega C}{\frac{1}{R}} = \omega CR.$$

Although the above formulae indicate that the Q of either component is proportional to frequency, in practice this is not the case for the following reasons :-

- (i) The representation adopted is approximate only and is not accurate at very high frequencies. For example, the shunt reactance of the self-capacitance of the coil reduces the effective inductance and eventually the coil resonates with its self-capacitance as a parallel resonant circuit.
- (ii) The resistance of a coil is not constant; but, owing to skin effect, which causes the current to be forced more and more towards the surface of the conductor, the resistance increases as the frequency rises.

For these and other reasons there is a maximum value for the Q - factor of any coil (or condenser) occurring at a definite frequency; (Fig. 15). Practical values for the Q - factor of a coil for use at radio frequencies of several megacycles per second are from 50 - 200. The higher the value of Q the narrower is the frequency band over which it is maintained.

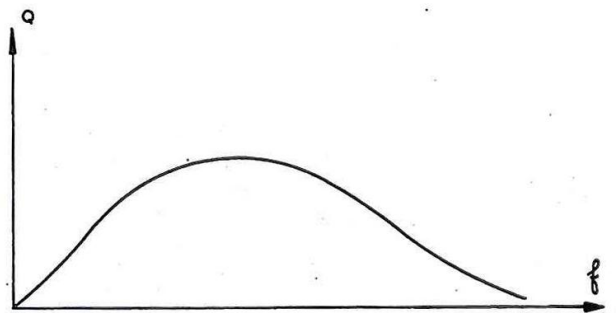


Fig. 15 - Variation of the Q -factor of a coil with frequency.

The Q - factor for an air-spaced condenser is very large compared with that of a coil. For a dielectric other than air or a vacuum the losses in the condenser are due almost entirely to the dielectric.

However, Q is still usually much greater for such a condenser than for the coils used at the same frequency, and since

$$Q = |\tan \phi| \quad \text{and is large, } \phi \doteq 90^\circ \text{ and } |\tan \phi| \doteq \frac{1}{\cos \phi}.$$

Hence the Q - factor of a condenser with a solid dielectric is approximately equal to $\frac{1}{F}$, where F is the power factor of the dielectric.

This may have a value from 100 to 5000.

17. Equivalent Series and Parallel Networks

A two-terminal network may be represented at a single frequency by :-

- (i) an equivalent series combination, or
- (ii) an equivalent parallel combination,

of simple components, as described in Secs. 13 and 14. In either case the impedance of the equivalent network must be the same as that of the actual network, having the same magnitude and phase angle.

Suppose z is the network impedance, of magnitude Z and phase angle ϕ , and y the admittance, of magnitude Y . Let the equivalent series combination consist of R_s and X_s in series, and the equivalent parallel combination of G_p and B_p in parallel. The following relations then hold :-

$$z = \frac{1}{y};$$

$$Z = \frac{1}{Y};$$

$$R_s = Z \cos \phi,$$

$$G_p = Y \cos \phi,$$

$$\text{Hence } R_s G_p = \cos^2 \phi$$

$$\text{or } R_s = \frac{R_p}{\cos^2 \phi}, \text{ (where } R_p = \frac{1}{G_p} \text{)}.$$

$$\text{Also, since } X_s = R_s \tan \phi$$

$$\text{and } B_p = -G_p \tan \phi,$$

$$\begin{aligned} X_s B_p &= -\tan^2 \phi \cos^2 \phi \\ &= -\sin^2 \phi \end{aligned}$$

$$\text{Hence } X_s = X_p \sin^2 \phi, \text{ (where } X_p = \frac{-1}{B_p} \text{)}$$

$$\text{The relations } R_s = R_p \cos^2 \phi,$$

$$\text{and } X_s = X_p \sin^2 \phi$$

make possible a rapid conversion from a series to a parallel arrangement or vice versa.

When the network consists of a single coil or condenser of large Q we may write $\tan \phi = Q$ where $Q \gg 1$.

$$\text{Then } \sin^2 \phi = \frac{Q^2}{1 + Q^2} \doteq 1$$

$$\text{and } \cos^2 \phi = \frac{1}{1 + Q^2} \doteq \frac{1}{Q^2}.$$

The equations relating series and parallel components may then be written

$$R_p \doteq Q^2 R_s$$

$$\text{and } X_p \doteq X_s.$$

18. FREQUENCY CHARACTERISTICS OF TWO-TERMINAL NETWORKS

Certain properties of simple linear networks with small losses can easily be deduced if the presence of resistive components is neglected, and a first approximation to the behaviour of the network can thereby be obtained. The impedance of any of the remaining elements, or combination of these elements, is then always reactive (or, in particular cases, zero or infinite).

For example, the reactance X_L of an ideal coil is ωL ,

and this is directly proportional to frequency; (Fig. 16(a)). The reactance X_C of an ideal condenser is $\frac{1}{\omega C}$ and is inversely

proportional to frequency, so that a graph showing the variation of X_C with f is a rectangular hyperbola; (Fig. 16(b)). The variation of the resultant reactance of these two in series is found by adding the ordinates of the curves (a) and (b), and is shown at (c).

For a certain value of f given by f_r , where

$$\omega_r L = \frac{1}{\omega_r C},$$

i.e.

$$\omega_r = \frac{1}{\sqrt{LC}},$$

so that

$$f_r = \frac{1}{2\pi\sqrt{LC}},$$

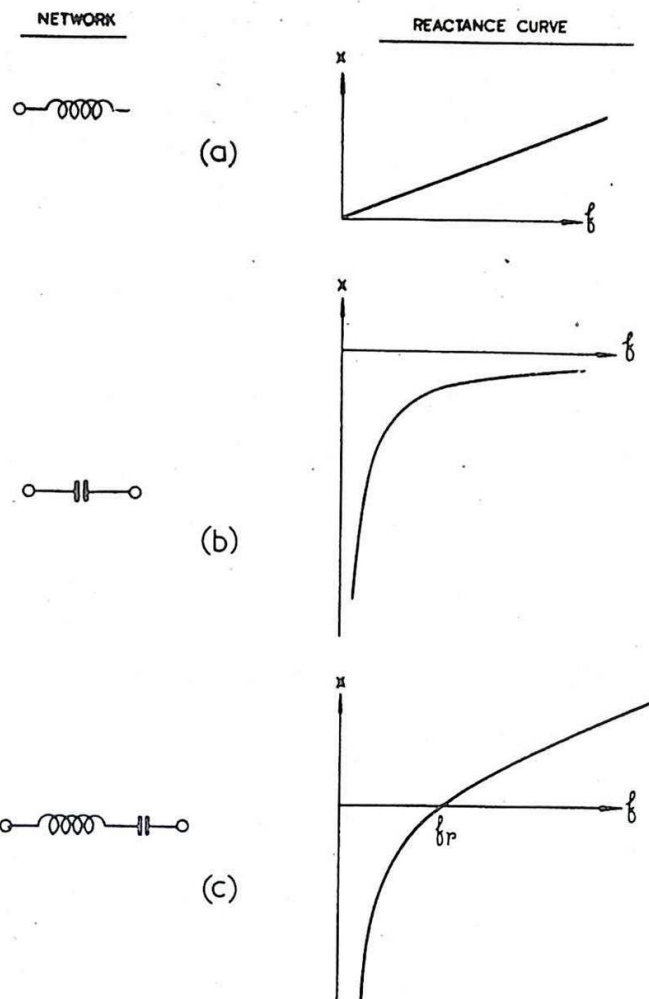


Fig. 16 - Reactance curves of simple non-resistive networks.

the resultant reactance of L and C in series is zero. This frequency is called the Resonant Frequency of the circuit. For lower frequencies this circuit, formed of ideal reactive components, is purely capacitive, and for higher frequencies, purely inductive.

For parallel circuits it is more convenient to use susceptances instead of reactances. The susceptance B_L of an ideal coil and B_C of an ideal condenser vary with frequency as illustrated in Fig. 17(a) and (b) respectively, and the resultant susceptance of the two in parallel is formed by adding the ordinates of the curves (a) and (b). This resultant is shown at (c). As for the reactance curve of the series circuit, this curve has a position of zero susceptance at the frequency

$$f = \frac{1}{2\pi\sqrt{LC}}$$

called the resonant frequency of the parallel circuit. For lower frequencies this circuit is purely inductive, and for higher frequencies, purely capacitive.

If the susceptance of any combination of L and C is known the reactance may be obtained by taking the reciprocal of the susceptance and changing its sign, since for a purely reactive impedance $X = \frac{-1}{B}$. The

converse process also holds. Equivalent reactance and susceptance curves obtained in this way for the series and parallel combinations already described are shown in Figs. 18(a) and (b), together with similar curves for more complicated networks, at (c) and (d).

An important property of these reactance and susceptance curves is that the slope is always positive. The reason for this is clear for the simple expressions $X = \omega L$ or $\frac{-1}{\omega C}$,

$B = \omega C$ or $\frac{-1}{\omega L}$, since

in each case dif-

ferentiation with respect to ω or f yields a positive derivative. The general case may easily be shown by induction. If X_1 is the reactance of any combination of L and C, and either L_1 or C_1 is added in series with X_1 then if $\frac{dX_1}{d\omega}$ is positive so also is $\frac{d}{d\omega} (X_1 + \omega L_1)$ or $\frac{d}{d\omega}$

$(X_1 - \frac{1}{\omega C_1})$. In fact, each additional series reactance increases

the slope at that frequency. A similar proposition is true if a susceptance is added in parallel with B_1 . Hence, since $\frac{dX}{d\omega}$ and

$\frac{dB}{d\omega}$ are always positive for the simplest networks, namely those con-

sisting of a single reactance, they are always positive for any series or parallel combination of such networks.

This fact makes the sketching of reactance and susceptance curves relatively simple. One is seldom interested in the exact variation of reactance or susceptance with frequency. It is usually sufficient in the first instance to plot approximately the frequencies at which the ordinate of the graph considered is zero or infinite, i.e.,

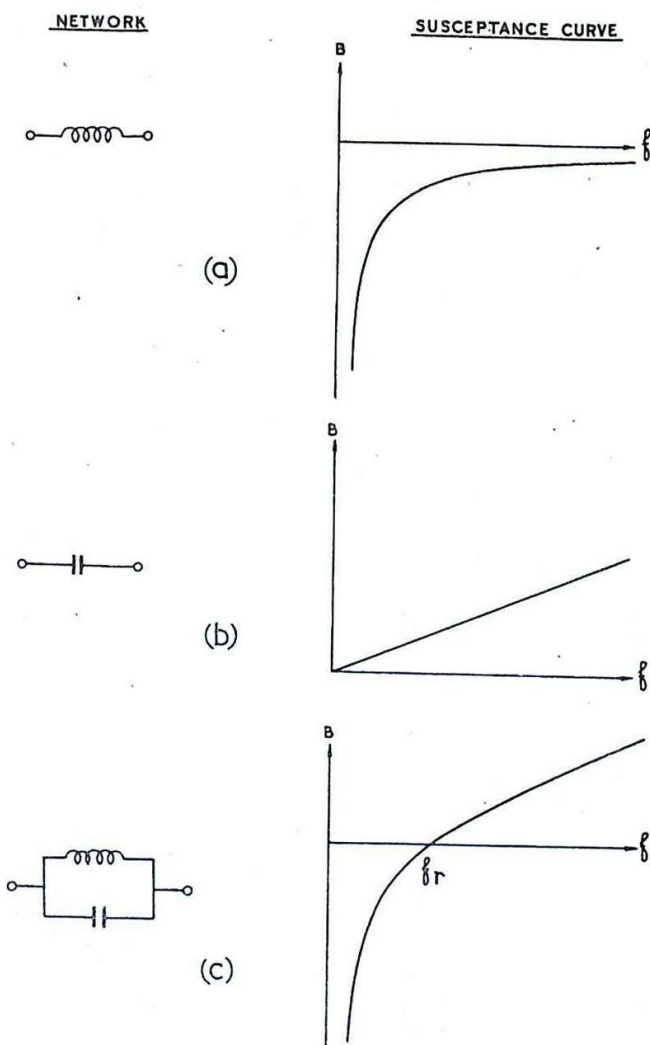


Fig. 17 - Susceptance curves of simple non-resistive networks.

the frequencies of series and parallel resonance. The general shape of the graph can immediately be sketched approximately once these points are determined.

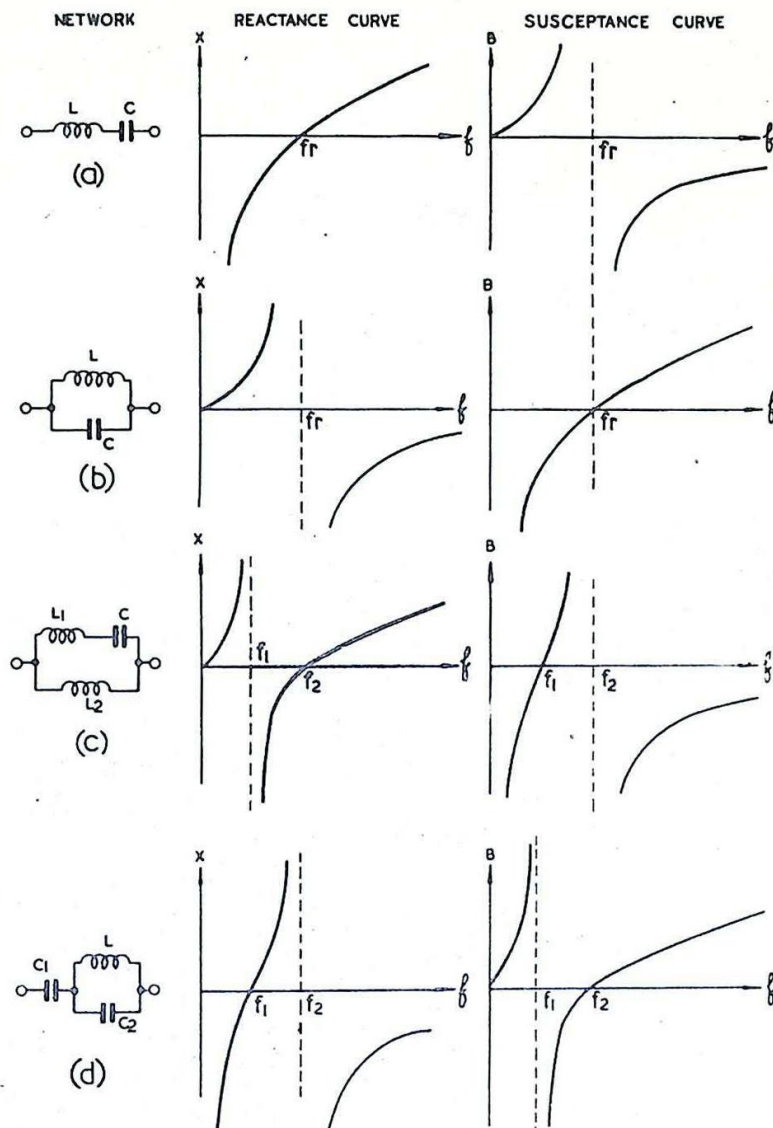


Fig. 18 - Reactance and susceptance curves of various non-resistive networks.

It has been possible to plot the variation of impedance or admittance, showing the variation both of magnitude and phase angle ($\pm 90^\circ$) on a single graph in Figs. 16 - 18 because the components were chosen as ideally reactive, having no resistance. In practice all components have resistance, and for accurate representation more than one two-dimensional cartesian diagram is necessary to represent the variation of magnitude and phase angle with frequency.

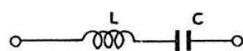
Fig. 19(a) shows the result of converting the reactance diagram of Fig. 16(c) for the Series Resonant Circuit, consisting of L and C in series, to a diagram representing the magnitude of the impedance. Since $R = 0$ and $Z = |X|$, the Z-curve is obtained by changing the sign of X in Fig. 16(c) where this is negative.

The phase variation is also shown.

We now consider the effect of inserting in series with L and C a resistance R, sufficiently small so that at the resonant frequency the reactance X_L (which is then equal to $-X_C$) is much greater than the resistance R.

The effect of R is appreciable only near resonance, since then the positive and negative reactances cancel each other, so that at the resonant frequency the impedance is R. For only small deviations from resonance the resultant reactance is much greater than R and the effect of R is small. This is illustrated in Fig. 19(b).

A more precise account of the behaviour of the circuit in the neighbourhood of the resonant frequency is given by the Universal Resonance Curve; (Sec. 19).



(a)



(b)

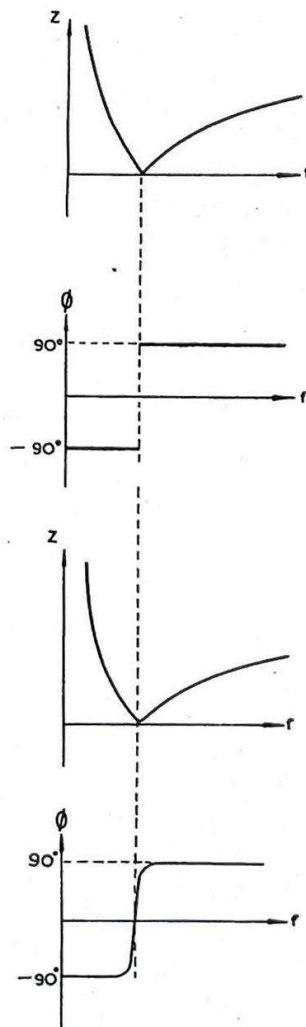


Fig. 19 - Effect of small series resistance on series L-C circuit.

A corresponding modification is needed to the diagram representing the magnitude of the admittance. Where the magnitude of the impedance is small, but not zero, as in Fig. 19(b), that of the admittance is large, but not infinite. Conversely, zeros in the admittance curve for resistance-less networks are replaced by finite admittances of small magnitude, corresponding to impedances of large, but not infinite magnitudes.

A similar procedure usually enables the behaviour of more complicated networks to be assessed qualitatively with little detailed analysis. The modifications necessary to the impedance curves of purely reactive networks to allow for slight losses usually consist of replacing zeros by small, and infinities by large, impedances and adapting the neighbouring portions of the diagrams accordingly. The resonant frequencies are little changed provided the losses are sufficiently small. (The procedure may be invalid if zeros and infinities in the reactance or susceptance graphs are very close together).

19. RESONANCE

It follows from Sec. 17, (Fig. 13(c)) that at a certain frequency the total reactance of L and C in series is zero. Above this frequency the circuit is inductive ($\phi = +90^\circ$) and below, capacitive ($\phi = -90^\circ$). Similarly, Fig. 14(c) shows that at some frequency the total susceptance of L and C in parallel is zero. Above this frequency the circuit is capacitive ($\phi = -90^\circ$) and below, inductive ($\phi = +90^\circ$). In both cases the resonant frequency is given by the relation

$$\omega_r L = -\frac{1}{\omega_r C}, \text{ so that } f_r = \frac{1}{2\pi\sqrt{LC}}$$

These results are modified by the presence of resistance, some resistance being unavoidable, notably the series resistance of the coil.

For the series (acceptor) circuit, if R is the total series resistance of the circuit, the impedance can be written

$$z = R + j\left(\omega L - \frac{1}{\omega C}\right)$$

and at resonance the reactive component is zero so that $z = R$. The reactive component grows rapidly and becomes large compared with R for very small deviations from resonance provided the ratio

$$\frac{\text{the magnitude of the reactance of either kind at resonance}}{\text{the total series resistance of the circuit}}$$

is large. This ratio is called the Q - factor of the circuit. It follows that

$$Q = \frac{\omega_r L}{R} = \frac{1}{\omega_r C R} = \frac{1}{R} \sqrt{\frac{L}{C}}$$

Since any equivalent series resistance associated with the condenser is usually negligible by comparison with the series resistance of the coil, in most series circuits used for HF work R is the resistance of the coil, so that the Q - factor of the circuit is the same as the Q - factor of the coil at the resonant frequency of the circuit.

We may write

$$\omega L = R \cdot \frac{\omega}{\omega_r} \cdot \frac{\omega_r L}{R} = RQ \frac{\omega}{\omega_r}$$

$$\text{and } \frac{1}{\omega C} = R \cdot \frac{\omega_r}{\omega} \cdot \frac{1}{\omega_r C R} = RQ \frac{\omega_r}{\omega}$$

$$\text{Hence } z = R + j RQ \left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega} \right)$$

$$\text{If we put } f = f_r (1 + \delta)$$

$$\text{then } \frac{\omega}{\omega_r} = \frac{f}{f_r} = 1 + \delta,$$

$$\text{and } \frac{\omega_r}{\omega} = \frac{1}{1 + \delta} \approx 1 - \delta \text{ provided } \delta \text{ is small.}$$

Hence $\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega} \doteq 2\delta$

Thus $z = R(1 + j2Q\delta)$.

As is shown in Sec. 14,

$$Z = R \sec \phi ;$$

also $\tan \phi = \frac{X}{R}$,

$$\text{but } z = R + jX$$

$$= R + jR \tan \phi$$

$$= R(1 + j \tan \phi).$$

Comparing this with

$$z = R(1 + j2Q\delta) \text{ above, we see that}$$

$$\tan \phi = 2Q\delta.$$

The approximations involved in the above analysis depend on δ being small. This is true for quite large values of Q provided Q is large. The variation of ϕ and $\cos \phi$ with $Q\delta$ is shown in Fig. 20. This graph is known as the Universal Resonance Curve. It may be used to represent the response of any resonant circuit in the neighbourhood of resonance provided Q is known and is large.

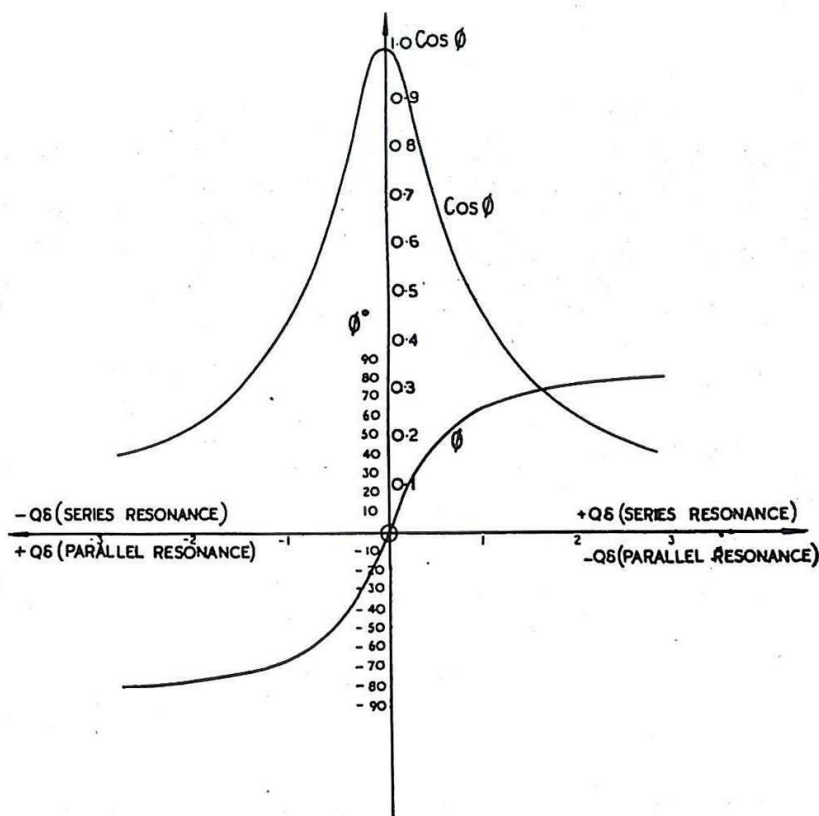


Fig. 20 - Universal resonance curve.

The analysis of the parallel (rejector) circuit of Fig. 15(c) when the resistance of the coil is taken into account is more complicated than that given above for the series resonant circuit. (See :- BR 229 Paras. 316 and 317; AP 1093 Chap. V Paras. 48 - 53)

A simple approximate treatment can be obtained using the results of Sec. 17. For any angular frequency ω we may replace the series combination of L and R by the parallel combination of L_p and R_D as shown in Fig. 21(a), provided

$$R = R_D \cos^2 \phi_L$$

and $L = L_p \sin^2 \phi_L$, where $\tan \phi_L = \frac{\omega L}{R} = \frac{R_D}{\omega L_p}$

The analysis of the network formed by R_D , L_p and C in parallel (Fig. 21(b)) is then very similar to that of the network containing R , L and C in series, and the resonant frequency is given by

$$f_r = \frac{1}{2\pi\sqrt{L_p C}}$$

At resonance the susceptance of C is exactly cancelled by that of L_p and the resultant impedance is the resistance R_D , called the Dynamic Resistance of the circuit. G_D , the reciprocal of R_D , is the Dynamic Conductance. The Q - factor of the circuit is defined as the ratio

$$\frac{\text{the magnitude of the susceptance of either kind at resonance}}{\text{the dynamic conductance}}$$

and therefore

$$Q = \frac{\omega_r C}{G_D} = \frac{\omega_r L_p}{C_D} = \omega_r C R_D = \frac{R_D}{\omega_r L_p}$$

and this is equal to $\frac{\omega_r L}{R} = \tan \phi_L$ = the Q - factor of the coil at the resonant frequency, ϕ_L being the phase angle of the coil impedance at this frequency.

Provided Q is large so that $Q^2 \gg 1$, we may put $L_p \doteq L$ and $R_D \doteq Q^2 R$. Also, the substitution of the parallel network of Fig. 21(a) for the series arrangement may be assumed to hold for the same values of L_p and R_D in the immediate neighbourhood of resonance.

The following relations then hold :-

$$Q \doteq \frac{1}{\omega_r C R} \doteq \frac{1}{R} \sqrt{\frac{L}{C}} \doteq R_D \sqrt{\frac{C}{L}} ;$$

$$f_r \doteq \frac{1}{2\pi\sqrt{LC}}$$

$$\text{In all cases } R_D = \frac{L}{C R} .$$

The exact expression for f_r can be obtained by substituting for L_p the exact relation

$$\begin{aligned} L_p &= L \operatorname{cosec}^2 \phi_L \\ &= L \left(1 + \frac{R^2}{\omega_r^2 L^2} \right). \end{aligned}$$

Since $\omega_r^2 L_p C = 1$, this gives

$$\omega_r^2 L \left(1 + \frac{R^2}{\omega_r^2 L^2} \right) C = 1$$

$$\text{or } \omega_r^2 LC = 1 - \frac{QR^2}{L}$$

$$\text{Hence } \omega_r = \frac{1}{\sqrt{LC}} \sqrt{1 - \frac{QR^2}{L}}$$

$$= \frac{1}{\sqrt{LC}} \sqrt{1 - \frac{1}{Q^2}}.$$

Unless $\frac{QR^2}{L} < 1$ there is no frequency at which the susceptance is zero, and the network is capacitive at all frequencies; (see Sec. 21).

Effect of Parallel Damping on Rejector Circuit

The value of Q for the undamped circuit (i.e. containing no additional damping other than the series resistance R) may be denoted by ${}_0Q$, and the corresponding dynamic resistance by ${}_0R_D$. If additional resistance R' is placed in parallel with the network, R_D is correspondingly modified, and may be denoted by ${}_1R_D$:-

$$\frac{1}{{}_1R_D} = \frac{1}{{}_0R_D} + \frac{1}{R'}; \quad (\text{Fig. 21 (c)}).$$

The Q - factor of the resultant circuit is modified according to the relation

$$\frac{{}_0Q}{{}_0R_D} = \frac{{}_1Q}{{}_1R_D}, \quad \text{so that for}$$

parallel damping, $Q \propto R_D$, other parameters remaining unchanged.

NOTE:- EQUIVALENCES
DENOTED BY \equiv ARE VALID
ONLY IN THE NEIGHBOURHOOD
OF ONE FREQUENCY, f_r .

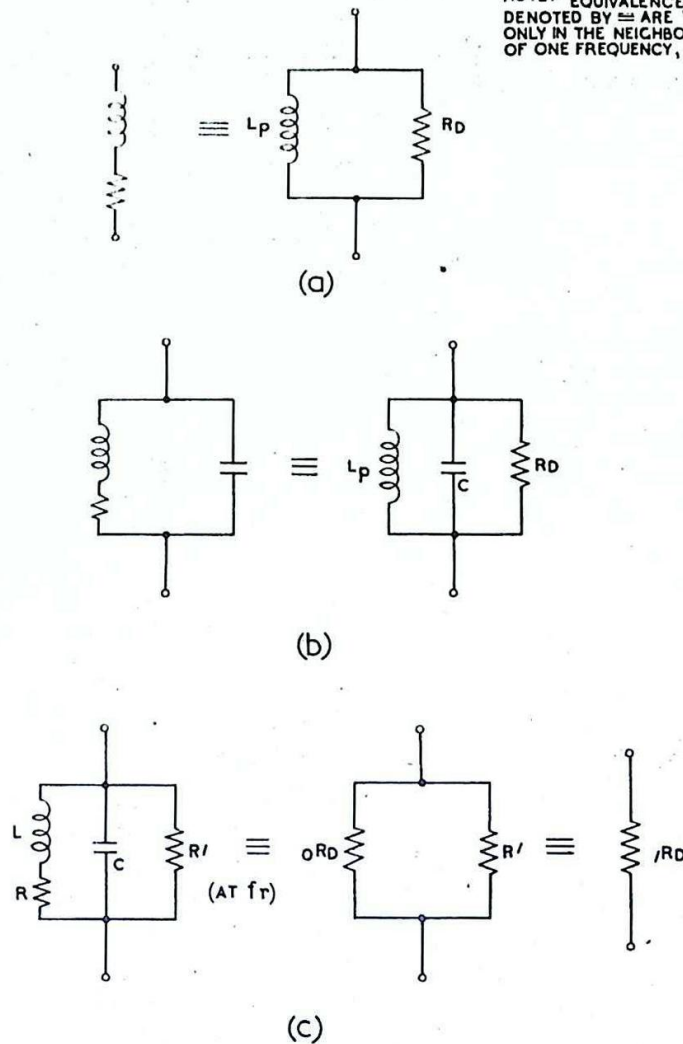


Fig. 21 - Rejactor circuit.

Use of Universal Resonance Curve

Provided Q is large the universal resonance curve gives accurately the variation of impedance (or admittance) of a series or parallel resonant circuit in the neighbourhood of resonance. For example, consider an acceptor circuit fed by an EMF of constant amplitude \hat{v}_i and variable frequency f ; (Fig. 22 (a)). The resonant frequency is f_r , and Q and R are known. The current at resonance i.e., when the input EMF is of frequency f_r , is of amplitude $\hat{i}_r = \frac{\hat{v}_i}{R}$.

At other frequencies the amplitude of the current is :-

$$\text{thus } \hat{i} = \frac{\hat{v}_i}{Z} = \frac{\hat{v}_i}{R \sec \phi} = \frac{\hat{v}_i}{R} \cos \phi = \hat{i}_r \cos \phi.$$

The universal resonance curve gives the variation of $\cos \phi$, but plotted against $Q\delta$. However, since Q and f_r are known f can be determined from the relation

$$\delta = \frac{f - f_r}{f_r}$$

so that $f = f_r \left(1 + \frac{Q\delta}{Q} \right) ;$

and for each value of $Q\delta$ the corresponding frequency f can be found.

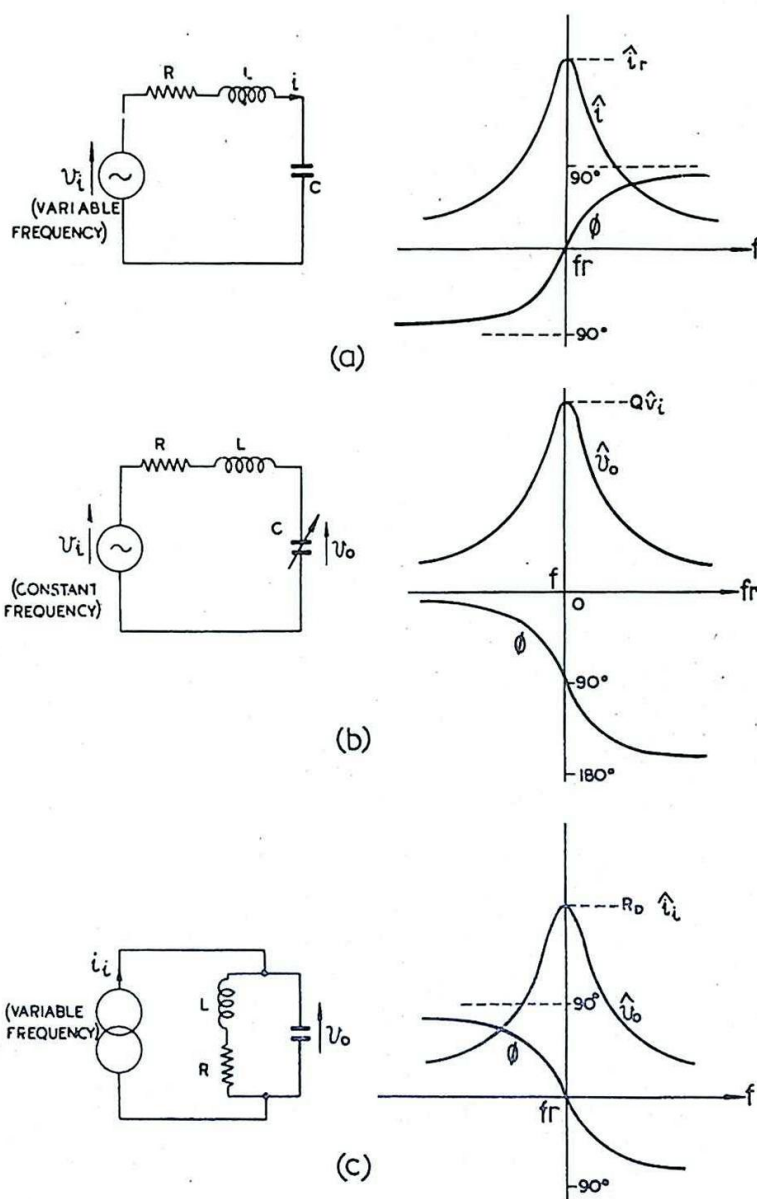


Fig. 22 - Use of universal resonance curve.

The response can then be represented as in Fig. 22(b).

When the current is divided by \hat{v}_i the curve shows the variation of the magnitude of the admittance.

Alternatively, the EMF may be of constant amplitude \hat{v}_i and frequency f , while one of the reactive components is varied as illustrated in Fig. 22(b). Provided Q is large and the series resistance of the circuit is constant, the universal resonance curve may be used to represent the variation of the voltage across either reactive component as the resonant frequency is varied. E.g., the

capacitance of the condenser may be varied and the movement calibrated in resonant frequencies. For a given Q , f_r may be determined for each value of $Q\delta$ from the relation

$$f_r = \frac{f}{1 + \frac{Q\delta}{Q}} = f(1 - \frac{Q\delta}{Q}).$$

The output voltage at resonance is $Q\hat{v}_i$, so that the variation is as shown in Fig. 22(b). The ϕ -variation of v_o is obtained by first plotting the variation of phase of the circuit current; for $f_r > f$, i.e., below resonance, the circuit is capacitive, and above resonance, for $f_r < f$, inductive. The required phase-variation of the output voltage is similar, but delayed by 90° on the phase of the current, and hence appears as shown in the figure.

Fig. 22(c) illustrates the use of the universal resonance curve for a rejector circuit fed by an ideal current generator of constant amplitude \hat{i}_i but varying frequency f . The output voltage has amplitude \hat{v}_o where

$$\begin{aligned}\hat{v}_o &= \hat{Z} i_i \\ &= \hat{i}_i R_D \cos \phi.\end{aligned}$$

The method of determining the frequency scale is the same as for the series circuit discussed first. When the ordinate is divided by \hat{i}_i the curve becomes one of impedance.

20. COUPLED CIRCUITS

The theory of simple coupled circuits is considered in Service Reference Books as follows:-

BR 229 Paras. 334 - 337.

26/Manuals/1577 Sec. 10 (and Chap. II app.B.)

AP 1093 Chap. VI. Paras. 23 - 31.

A few additional remarks are made here with a view to presenting some of the results in the form in which they are used in the remaining chapters.

Two simple series networks coupled by means of their mutual inductance are illustrated in Fig. 23. The primary circuit is considered to be fed by an ideal voltage generator of EMF v_G . The mutual inductance between the circuits is M ; the total secondary impedance (i.e. the impedance of the secondary circuit when $M = 0$) is:-

$$z_2 = R_2 + j(\omega L_2 - \frac{1}{\omega C_2}).$$

The total primary impedance, i.e. the impedance of the primary circuit when $M = 0$, is

$$z_1 = R_1 + j(\omega L_1 - \frac{1}{\omega C_1}),$$

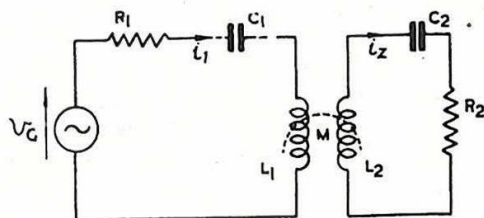


Fig. 23 - Circuits coupled by mutual inductance.

but it will be shown that this is not the impedance presented to the generator.

The EMF induced in the secondary circuit due to i_1 in the primary is

$$\begin{aligned} v_2 &= M \frac{di_1}{dt} \\ &= j\omega M i_1. \end{aligned}$$

This causes a current i_2 to flow in the secondary circuit given by

$$i_2 = \frac{v_2}{z_2} = \frac{j\omega M i_1}{z_2}.$$

Due to this current an EMF is induced in the primary circuit given by

$$\begin{aligned} v_1 &= M \frac{di_2}{dt} \\ &= j^2 \frac{\omega^2 M^2}{z_2} i_1 \\ &= - \frac{\omega^2 M^2}{z_2} i_1. \end{aligned}$$

The net EMF in the primary circuit is therefore $v_G + v_1$, so that

$$v_G - \frac{\omega^2 M^2}{z_2} i_1 = z_1 i_1$$

and therefore the impedance presented to the generator is

$$z = \frac{v_G}{i_1} = z_1 + \frac{\omega^2 M^2}{z_2}.$$

The second term on the right hand side of this equation is called the Reflected Impedance of the secondary circuit.

21. BROAD-BANDING

The principle of Broad-banding is one of wide application in radar circuits. In its extreme form the requirement is to design a network which will pass uniformly signals of all frequencies within a wide band and which will reject all others; but for most purposes a rough approach to this ideal is satisfactory. Band-pass circuits are dealt with elsewhere, and this section is restricted to the consideration of a network which is essentially a simple type of low-pass filter: i.e. it gives a response which is reasonably uniform for low frequency signals, but which eventually falls off as the frequency is raised.

The impedance of an ideal resistance is constant at all frequencies. In practice this cannot be achieved, particularly because of unavoidable shunt capacitance, which reduces impedance at high frequencies. This is illustrated in Fig. 24(a). For low frequencies,

$$Z \approx R;$$

but for higher frequencies, at which $\frac{1}{\omega C}$ is small enough to be comparable with R , the magnitude of the impedance is given by

$$Z = R \cos \phi,$$

and this becomes very small as ϕ approaches -90° , i.e. for $\omega CR \gg 1$.

If a coil is inserted in series with R , as shown at (b), the increase in the impedance of the series arm as the frequency rises partially offsets the fall in the impedance of the condenser, and the frequency response is improved.

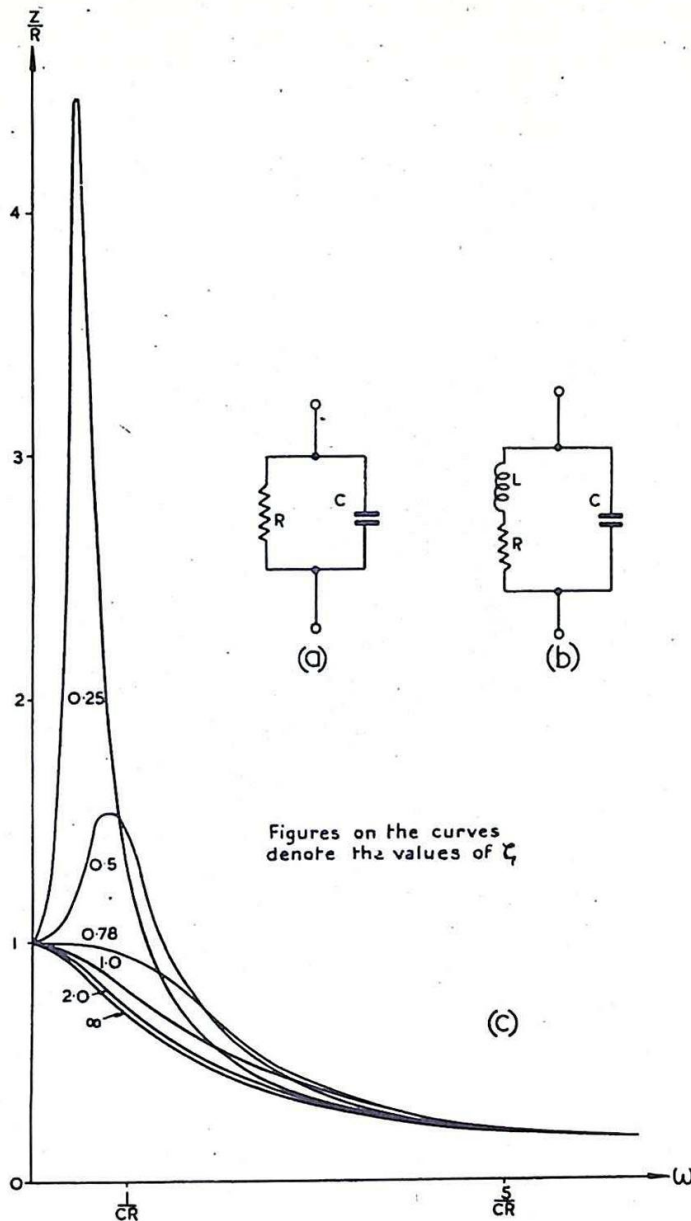


Fig. 24 - Broadbanding; use of series inductance to counteract shunt capacitance.

It is not practicable to use the universal resonance curve for this heavily damped circuit, since the Q - factor is very small. A more useful criterion is the Damping Ratio, which for this circuit is given by

$$\zeta = \frac{R}{2} \sqrt{\frac{C}{L}}$$

$$(\text{Compare } Q \doteq \frac{1}{R} \sqrt{\frac{L}{C}})$$

when Q is large).

For $\zeta > \frac{1}{2}$ there is no "resonant frequency" i.e. no frequency at which the phase-angle of the network is zero, but the network is capacitive at all frequencies.

The variation of the magnitude of the impedance with frequency for different values of ζ is shown in Fig. 24(c). The magnitude of the impedance is maintained approximately equal to R for a relatively wide frequency range if ζ is made about 0.78. For smaller values of ζ , i.e., larger L , the network exhibits resonance properties, with peaks to the impedance-frequency diagram which may be undesirable. For larger values of ζ , i.e., smaller L , the bandwidth is reduced. Other slightly different values of ζ may be chosen for various reasons, e.g., maximum constancy of phase- or time-delay instead of maximum constancy of the magnitude of the impedance.

CHAPTER 2

RESPONSE OF LINEAR CIRCUIT ELEMENTS TO VOLTAGE PULSES

1. INTRODUCTION

In alternating current theory, it is normal to consider the results of applying sinusoidal voltages to circuits containing resistance, capacitance and inductance. This theory has common application to power and radio-communication systems. In radar and television systems it is equally common to apply to such circuits voltages which are by no means sinusoidal.

When the applied voltage is sinusoidal, the voltages produced across the individual components are also sinusoidal, differing only in phase and magnitude from the input. The relative phases and magnitudes depend on the frequency of the input and on the relative magnitudes of the components.

When the applied voltage is non-sinusoidal the voltages developed across the circuit elements are distorted versions of the input. This makes the circuit behaviour much more complicated, each type of input requiring individual investigation for different relative magnitudes of the circuit components.

This chapter will be devoted to a consideration of the results of applying non-sinusoidal voltages to simple circuits containing linear elements only.

INSTANTANEOUS APPLICATION OF CHANGES OF VOLTAGE TO CIRCUITS CONTAINING CAPACITANCE AND RESISTANCE

2. Instantaneous Application of Change of Voltage To a Series C-R Circuit

Suppose a sudden change of voltage as shown in Fig. 25 is applied to a circuit containing C and R in series. Let the input voltage v_i rise instantaneously from zero to \hat{v}_i at time $t = 0$, and thereafter be maintained indefinitely at this value. Denote the voltage developed across the condenser, of capacitance C, at any instant, by v_C and the voltage across the resistor, of resistance R by v_R . Simple theory (Admiralty Handbook of Wireless Telegraphy, BR229, para. 174, and AP 1093 Part II, Chap. VII, para. 59) shows that the voltage v_C rises exponentially as given by

$$v_C = v_i (1 - e^{-t/CR})$$

and as illustrated in Fig. 25. By Kirchhoff's law we know that the sum of the voltages across the capacitance and resistance must at all instants equal the applied voltage, i.e.,

$$v_i = v_R + v_C,$$

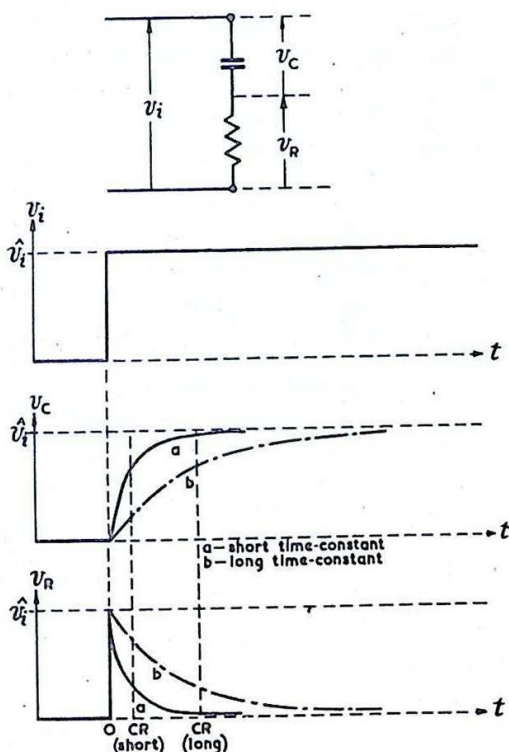


Fig. 25. - Response of a C-R circuit to an instantaneous rise of voltage.

so that the voltage v_R is given by

$$v_R = \hat{v}_i e^{-t/CR}$$

as shown in Fig. 26

The time-constant CR appropriate to these changes represents the time taken for the voltage across the resistance to fall to $\frac{1}{e} \hat{v}_i$ i.e., about one third of its original value. After a time equal to about $5CR$ has elapsed the voltage has fallen to less than 1% of its original value, and for most practical purposes the change may be considered complete.

If the sudden change of voltage is a decrease instead of an increase similar considerations apply, and the results are as shown in Fig. 26

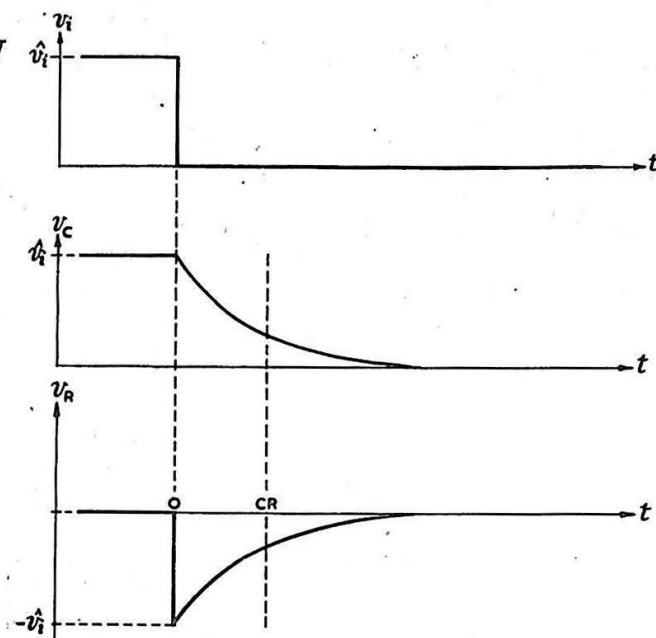


Fig. 26.- Response of a C-R circuit to an instantaneous fall of voltage.

3. Application of a Rectangular Pulse of Voltage to a Series C-R Circuit

We now consider the application of a voltage whose time variation is as shown in Fig. 27. We shall call this a Rectangular Pulse of Voltage. It is simplest to consider in turn firstly the effect of the sudden increase of voltage at the start of the pulse and secondly the effect of the sudden decrease at the end. We have already dealt with these two cases separately.

Assume first that the time-constant CR of the circuit is considerably less than the time of duration T of the pulse. Then the charging of the condenser is completed (in so far as an exponential rise is ever completed) before the discharge takes place, and the voltages developed across the condenser and resistor are as shown in Fig. 27. Two sharp narrow pulses of voltage of opposite sign are produced across the resistor, one at the start and one at the end of the applied pulse. The durations of these sharp pulses depend on the magnitude of the time-constant.

If the time-constant of the circuit is made variable (usually, for convenience, by changing the value of R) the width of these short duration pulses developed across the resistor can be controlled. In the ideal case considered the full voltage \hat{v}_i would be developed across the resistor however short the time-constant. In practice this is not so for reasons

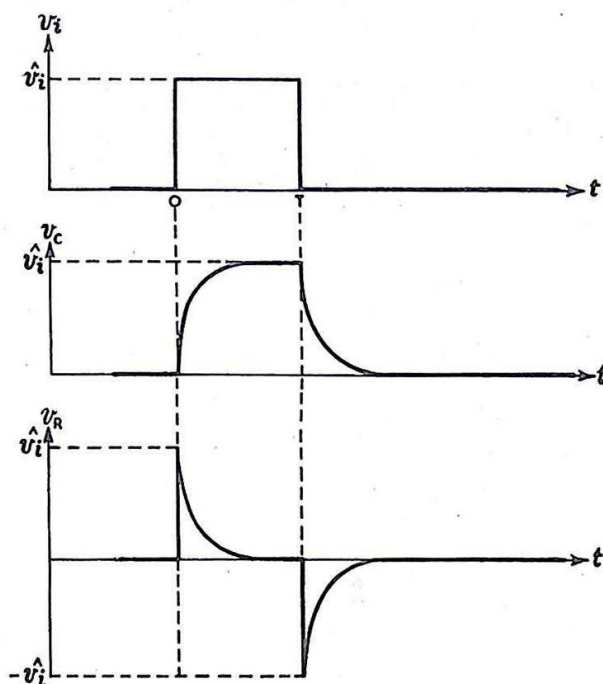


Fig. 27.- Response of a C-R circuit to a rectangular pulse. $CR \ll T$

discussed later (Secs. 5 and 12), and the amplitude of v_R tends to be further reduced the smaller the value of the time-constant.

Whilst the duration of the short pulses developed across the resistor depends on the time-constant of the circuit, it does not depend on the duration T of the input rectangular pulses. The time which elapses between the leading edges of the first and second pulses developed across the resistor is equal to T .

Suppose next that the time-constant of the circuit is very much greater than the duration of the applied pulse. In this case the voltage across the condenser rises only slightly before the input voltage drops and causes the condenser to discharge (Fig. 28). Consider the case when $CR = 10T$; then at time T (time at which input voltage drops) v_C is given by :-

$$\begin{aligned} v_C &= \hat{v}_i (1 - e^{-t/CR}) \\ &= \hat{v}_i (1 - e^{-1/10}) \\ &= \hat{v}_i (1 - 0.905) \\ &= 0.095\hat{v}_i \end{aligned}$$

After time T the condenser discharges from the value $0.095\hat{v}_i$ towards zero. The value of v_C after time $2T$ (time T after the instant at which the input voltage falls to zero) is given by :-

$$\begin{aligned} v_C &= 0.095\hat{v}_i e^{-t/CR} \quad (0.095\hat{v}_i = \text{potential difference at start of discharge}) \\ &= 0.095\hat{v}_i e^{-1/10} \\ &= 0.095\hat{v}_i \times 0.905 \\ &= 0.086\hat{v}_i \end{aligned}$$

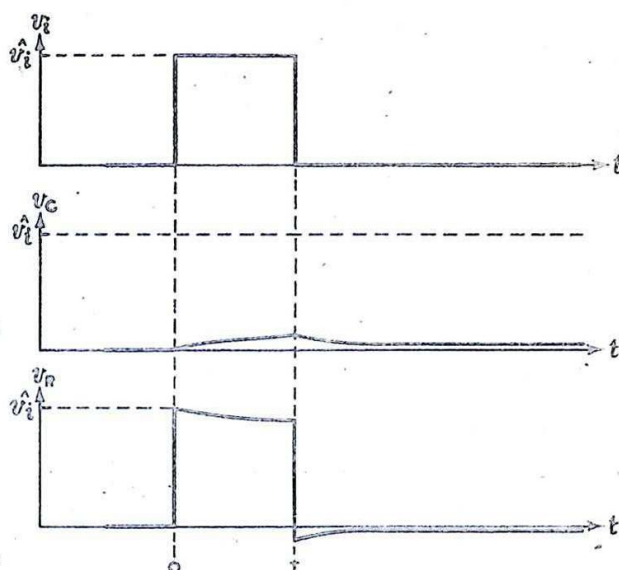


Fig. 28. - Response of a C-R circuit to a rectangular pulse. $CR \gg T$.

This difference between the rates of charging and of discharging of the condenser under the above conditions can be summed up as follows. The rate of charge or discharge is proportional to the voltage applied to the resistor. At the beginning of the charging period this voltage is \hat{v}_i , whilst at the beginning of the discharging period it is only $0.095\hat{v}_i$. Thus the rate of discharge is less than one-tenth the rate of charge. The voltage v_R across the resistor rises to a value \hat{v}_i at the onset of the rectangular input pulse, falls, at the same rate as v_C rises, for the duration T , and then drops instantaneously by an amount \hat{v}_i with the end of the input pulse. Finally v_R rises exponentially towards zero at the same rate as v_C falls. The longer the time-constant CR , the more faithfully does v_R reproduce the input voltage. Thus when $CR = 100T$ the drop in v_R throughout the duration T of the input pulse is only $0.01\hat{v}_i$ and the output pulse across the resistor is very little distorted compared with the input.

In radar systems it is common practice to apply a rectangular pulse to a C-R circuit such as that of Fig. 28 and to utilise the voltage developed across the resistor. The above considerations reveal the following points :-

- (i) The maximum value of the voltage developed across R is in all cases equal to the magnitude of the applied pulse.
- (ii) The shape of the output voltage is a close replica of the input if $CR \gg T$; but if $CR \ll T$ the input rectangular pulse is converted

into two narrow pulses of opposite sense to each other, whose durations depend on the value of the time-constant CR.

4. Application of a Succession of Rectangular Pulses of Voltage To a Series C-R Circuit

We now consider the case where a succession of rectangular pulses of voltage is applied to the circuit of Fig. 25. Suppose T_1 is the duration of each pulse and T_2 the interval between the pulses.

The general shapes of the voltages across C and across R are similar to those already described. However, the mean level of the voltage across the condenser tends to rise with each successive pulse; unless the time-constant of the circuit is so short compared with either T_2 or T_1 that the condenser can be considered to have discharged or charged completely during these periods.

Ultimately a Steady State is reached in which the charge acquired by the condenser during the time T_1 is equal to the

charge lost during the time T_2 . This means that the mean current flowing into the condenser through the resistor is zero when equilibrium has been reached, so that the mean value of v_R is also zero. The graph showing the voltage developed across the resistor is such that the area (voltage x time) enclosed above the zero (mean) voltage line during the time T_1 , is equal to the area enclosed below this line during the time T_2 , i.e., in the steady state the shaded areas A and B (Fig. 29) are equal.

Alternatively, a useful way of approaching the problem is to consider the succession of positive input pulses to be composed of a steady component of voltage and a purely alternating component of voltage.

The mean value \bar{v}_i of the input voltage is given by :-

$$\bar{v}_i = \frac{T_1}{T_1 + T_2} \hat{v}_i$$

The alternating component of course has a mean value of zero. If a steady voltage is applied to a condenser and resistor in series, the whole of the voltage appears ultimately across the condenser and, as we have just seen, the rise is exponential. In the present case therefore the mean voltage across the condenser increases exponentially with time-constant CR towards the value

$$\frac{T_1}{T_1 + T_2} \hat{v}_i$$

Ultimately the mean voltage across the resistance is zero.

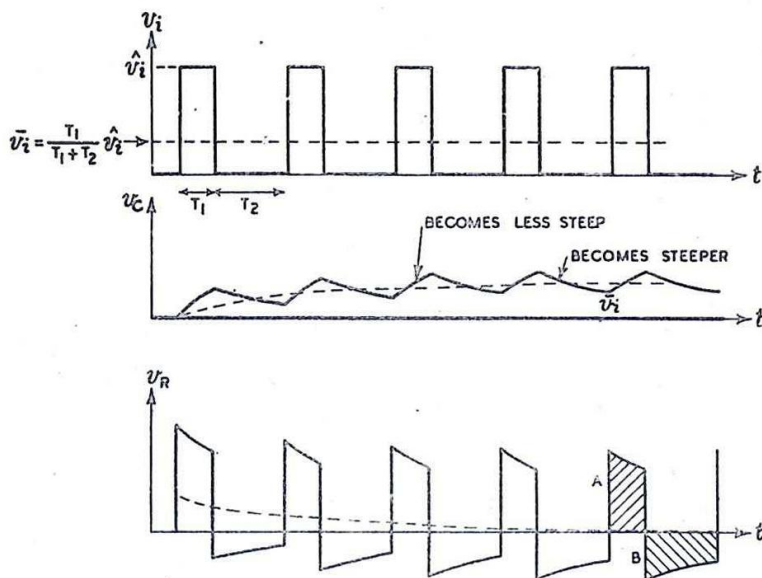


Fig.29.- Application of a succession of rectangular pulses to a C-R circuit. $CR \gg T_1$

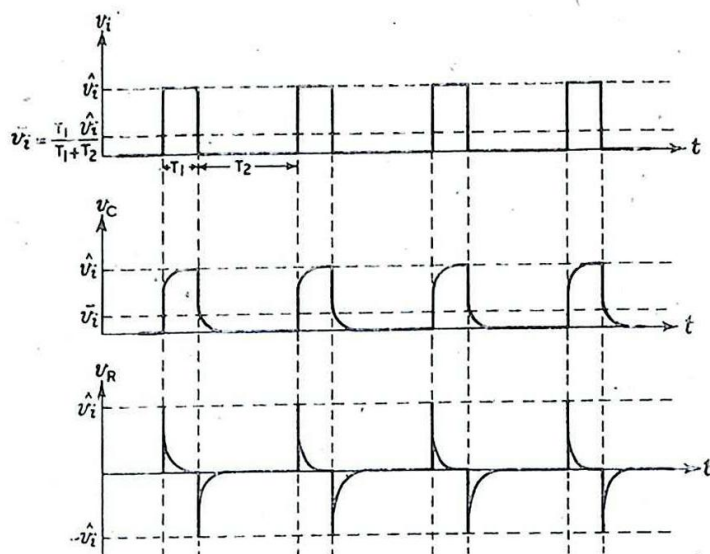


Fig.30.- Application of a succession of rectangular pulses to a C-R circuit. $CR \ll T_1$

It is therefore apparent that if $CR \gg T_1$ the output voltage across R is a replica of the input voltage across C in so far as the alternating portion of the input is concerned; but that the steady component of the input voltage variation is not reproduced. If it is required to reproduce this steady component, which is determined by the mean voltage level of the input, special measures have to be employed. This process of restoring the steady (or DC) component of voltage is often called DC Restoration (See Chap. 12 Sec 2).

If $T \gg T_1$, the input can be considered as a succession of negative-going pulses, and a similar argument holds. In this case the condition for faithful reproduction of the alternating component of the applied pulse in the voltage across the resistor is that $CR \gg T_2$.

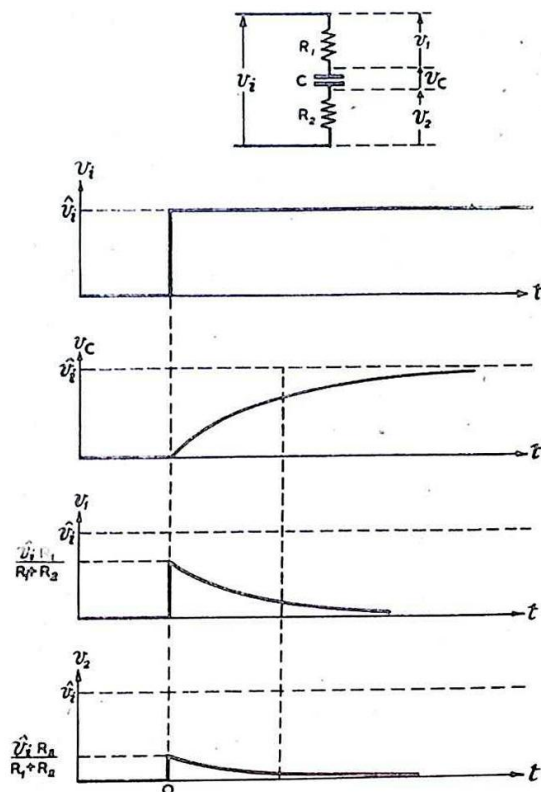
If CR is short compared with both T_1 and T_2 the steady state is rapidly approached, and this is illustrated in Fig. 30.

5. Instantaneous Application of a Change of Voltage to Circuits Containing More Than One Condenser or Resistor.

Fig.31 shows a circuit consisting of two resistors R_1 and R_2 in series with a condenser C . The sum of the voltages developed across R_1 and R_2 behaves exactly the same way as the voltage V_R across R in the circuit of Sec. 1-4, with

$$R = R_1 + R_2$$

Fig.31.- Response of a C-R circuit which contains more than one resistor.



The voltage developed across each resistor is proportional to its resistance, i.e.,

$$v_1 = \frac{R_1}{R_1 + R_2} v_R \text{ and } v_2 = \frac{R_2}{R_1 + R_2} v_R$$

The time-constant is

$$C (R_1 + R_2)$$

A circuit commonly encountered in radar is shown in Fig. 32. R_1 represents the output resistance of one amplifier, whilst C_2 is the input capacitance of the next stage. In general $C_1 \gg C_2$.

Unfortunately the general analysis of this circuit is too complicated for inclusion here, but if the output resistance of the pulse generator is made very small the circuit of Fig. 33 may be taken as a sufficiently close approximation.

At time $t = 0$ the input voltage rises almost instantaneously from zero to \hat{v}_i . (This rise is actually exponential, but because the generator output resistance is very small, the effective time-constant is negligible). This rise of voltage is developed across the two condensers in the inverse ratio of their capacitances. While the input voltage is maintained at \hat{v}_i , the condenser C_1 charges exponentially so that the voltage across it rises towards the value \hat{v}_i , and C_2 discharges exponentially towards zero. The time-constant of these variations can be shown to be $R_2 (C_1 + C_2)$.

It will be noticed that one effect of the condenser C_2 is to reduce the maximum value of the output voltage developed across the resistor R_2 . Thus, if the condenser C_1 is reduced in value in order that the time-constant may be small, i.e., so that the output voltage may consist of short duration pulses, the amplitude of these pulses is reduced.

INSTANTANEOUS APPLICATION OF CHANGES OF VOLTAGE TO CIRCUITS CONTAINING INDUCTANCE AND RESISTANCE

6. Instantaneous Application of a Change of Voltage to a Series L-R Circuit (Fig. 34).

Let the input voltage v_i rise instantaneously from zero to \hat{v}_i at time $t = 0$ and thereafter be maintained indefinitely at this value. Denote by v_L the voltage developed at any instant across the coil, of inductance L , and by v_R the voltage across the resistor, of resistance R .

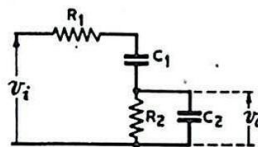


Fig. 32.- Practical form of C-R circuit.

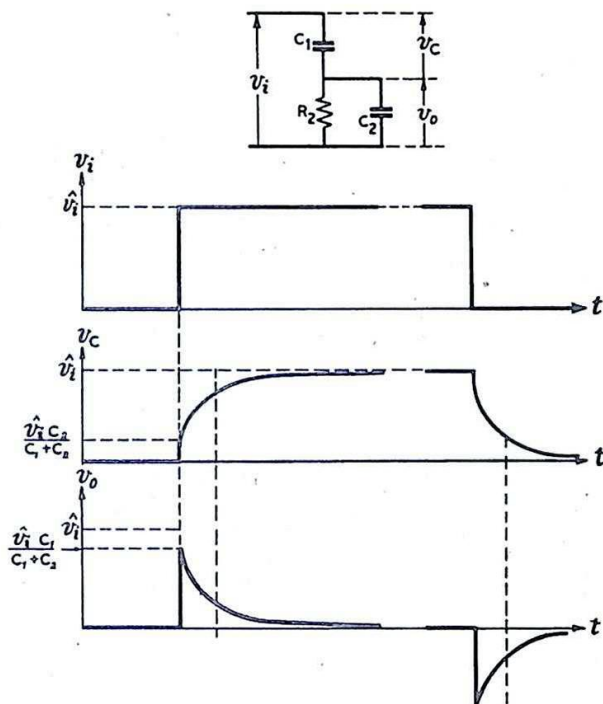


Fig. 33.- Response of practical C-R circuit (simplified).

Standard theory (BR229 Admiralty Handbook of Wireless Telegraphy, Vol. I, para. 158, and AP 1093 Part II Chap. II, para. 43) shows that the rise in current in the circuit is given by the expression :-

$$i = \frac{\hat{v}_i}{R} \left(1 - e^{-\frac{Rt}{L}} \right)$$

so that the voltage developed across the resistance R is

$$v_R = \hat{v}_i \left(1 - e^{-\frac{Rt}{L}} \right)$$

At any instant v_L may be determined by the relation

$$v_i = v_R + v_L$$

The time-constant is L/R .

Fig. 34 also shows the effect of reducing, instantaneously, the input voltage from a value \hat{v}_i to zero, assuming that the voltage across the coil and resistor have attained their final values before the change takes place. The full drop of voltage appears instantaneously across the coil. Subsequently the voltage across the resistor falls exponentially from \hat{v}_i towards zero, while that across the coil rises exponentially from a $-\hat{v}_i$ towards zero.

It should be noted that the nature of the voltage variation across the coil is similar to that across the resistor in the corresponding C-R circuit, whilst the voltage variation across the resistor in the present case is similar to that developed across the condenser.

7. Application of a Rectangular Pulse of Voltage to a Series L-R Circuit

Consideration similar to those of Sec. 3 apply to a circuit consisting of a coil L and resistor R in series. We shall consider separately the two cases (i) $\frac{L}{R} \ll T$ and (ii) $\frac{L}{R} \gg T$, T being the duration of the input pulse.

(i) $\frac{L}{R} \ll T$ (Fig. 34)

The voltage across the resistor rises exponentially to practically its full value \hat{v}_i during the interval T, while the voltage across the coil, after its initial instantaneous rise to \hat{v}_i , falls exponentially to zero. At time T the input voltage drops instantaneously to zero, so the voltage across the coil drops instantaneously to $-\hat{v}_i$. After time T the voltage across the coil rises exponentially from $-\hat{v}_i$ towards zero, while the voltage across the resistor falls exponentially from \hat{v}_i towards zero.

Thus voltage pulses of short duration are developed across the coil.

(ii) $\frac{L}{R} \gg T$

The voltage across the resistor rises exponentially by only a small amount before the end of the applied pulse; then the

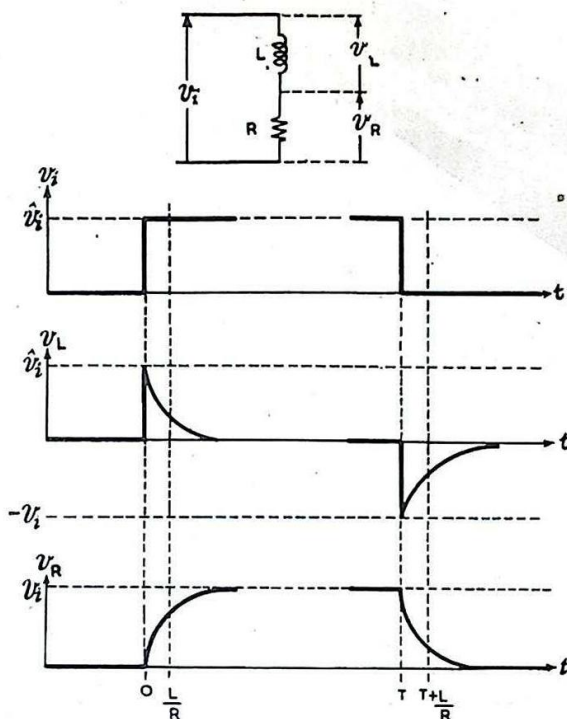


Fig. 34. - Response of L-R circuit to rectangular pulse. $L/R \ll T$

voltage across the resistor starts to fall exponentially. This behaviour is similar to that of v_C in Fig. 28. The voltage across the coil rises instantaneously from zero to a value \hat{v}_L at time $t = 0$, falls exponentially by a small amount during the interval $t = 0$ to $t = T$, and then drops instantaneously by an amount \hat{v}_L . Subsequently the voltage developed across the coil rises exponentially towards zero. The voltage developed across the coil is an almost undistorted reproduction of the input pulse.

8. Application Of a Succession Of Rectangular Pulses of Voltage to a Series L-C Circuit

As in the corresponding case of the C-R circuit (Fig. 29, Sec 4), the mean value of the input is given by :-

$$\bar{v}_i = \frac{T_1}{T_1 + T_2} \hat{v}_i$$

The series of pulses may be split up into a steady component equal to this mean value, and an alternating component whose mean value is zero. The steady component of voltage across the resistor ultimately attains the mean value \bar{v}_i , and approaches this value at a rate dependent upon the time-constant of the circuit. In the steady state only the alternating component, which represents the rectangular pulse shape, appears across the coil.

9. DISADVANTAGES OF A SERIES L-R CIRCUIT COMPARED WITH A SERIES C-R CIRCUIT

It is not usually convenient to employ L-R networks to provide time-constants of more than a few microseconds such as are commonly required in radar. Such a network is usually in series with some form of generator whose output resistance is of the order of a few thousand ohms, so that the size of coil required to give the desired $\frac{L}{R}$ ratio is

$$\frac{L}{R}$$

inconveniently large. The requirements can be met by the use of C-R components of more practicable values. Further, the inherent resistance of a condenser is usually negligible whereas that of a coil is always appreciable. Even when very short time-constant circuits are required, coils cannot be used to produce the effects described above because of their self-capacitance. Their use in ringing circuits, where self-capacitance is not necessarily deleterious, is described in Sec. 10.

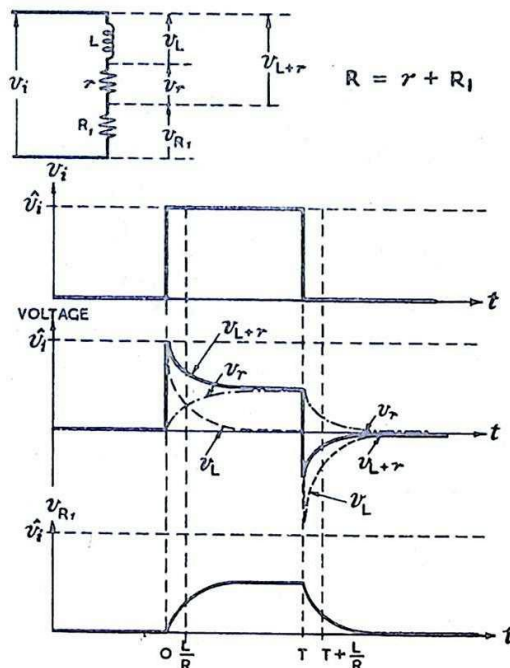


Fig. 35.- Modifications to response of L-R circuit due to coil resistance.

Fig. 35 shows the effect of coil resistance on the voltages across coil and resistor in a series L-R circuit. The voltage developed across the resistor is affected only in magnitude, but the coil voltage may be considerably modified as shown (compare v_{L+r} of Fig. 35 with v_L of Fig. 34).

INSTANTANEOUS APPLICATION OF CHANGES OF VOLTAGE TO CIRCUITS CONTAINING INDUCTANCE, CAPACITANCE AND RESISTANCE

10. Instantaneous Application of a Change of Voltage to An L-C-R Circuit

The consideration of free oscillations in a series L-C-R circuit (Fig.36(a)) is dealt with in many standard works (Admiralty Handbook of Wireless Telegraphy, BR229, para. 390, and AP 1093 Part II, Chap. VII, paras. 7-16). It is shown that the sudden application of a change of voltage to this kind of circuit results either in a free oscillation or in a smooth, non-oscillatory change, according to the relative values of the components.

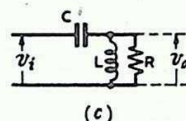
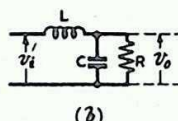
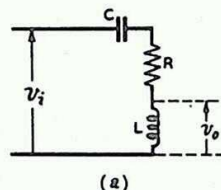


Fig.36.- L-C-R circuits.

In the case of a free oscillation the frequency depends mainly upon L and C and to a lesser extent upon R, and the amplitude decreases exponentially according to the amount of damping in the circuit.

The quantity $\zeta = \frac{R}{2} \sqrt{\frac{C}{L}}$ is called the Damping Ratio.

If $\zeta < 1$ the Circuit will perform a damped oscillation with a frequency f and a logarithmic decrement δ given by :-

$$f_o = \frac{1}{2\pi} \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

$$\delta = \frac{R}{2fL}$$

If the resistance is very small the frequency is approximately $\frac{1}{2\pi\sqrt{LC}}$. Such a circuit is called a Ringing Circuit.

If $\zeta = 1$ the output voltage of the circuit is just not oscillatory (Critical Damping).

If $\zeta > 1$ the output voltage takes the form of an exponential rise.

In many radar applications it is desirable to use a ringing circuit whose damping is slightly less than critical.

Ringing circuits are frequently used with the damping resistor R in parallel with either the coil or the condenser instead of in series, as shown in Figs.36(b) and (c). In these cases it can be shown that, neglecting the resistance of the coil, critical damping occurs when $R = \frac{1}{2} \sqrt{\frac{L}{C}}$, and that oscillations take place for values of R greater than this.

Fig. 37 shows the output voltage produced across the condenser in the circuit of Fig.36(b) when a change of voltage is instantaneously applied at the input terminals. Diagrams illustrate the cases $R = \frac{1}{2} \sqrt{\frac{L}{C}}$ (damped oscillatory), $R = \frac{1}{2} \sqrt{\frac{L}{C}}$ (critically damped) and $R = \frac{1}{4} \sqrt{\frac{L}{C}}$ (non-oscillatory). In all cases the delay-time before the output voltage reaches the value

of the input is proportional to \sqrt{LC} . The voltage across the coil may be obtained from the relation $v_i = v_L + v_C$.

11. INSTANTANEOUS APPLICATION OF A CHANGE OF VOLTAGE TO A DELAY NETWORK

A more detailed consideration of this problem is given in Chap. 4, where the wave-nature of the effect is dealt with. As an introduction, and for the sake of completeness, a brief mention is made here, considering the behaviour of the delay network as an extension of that of the ringing circuit.

Fig. 38 shows two-stage and six-stage networks composed of identical sections, such as are used to make up a pulse-forming Delay Line. In Sec. 10, illustrated by Figs. 36 & 37, we have already discussed the behaviour of the single-stage network, or ringing circuit. If we replace the resistance R of Fig. 36 (b) by another ringing circuit identical with the original one, we have the circuit of Fig. 38(a). We shall assume that this does not appreciably

affect the voltage v_1 produced across the first condenser when the input voltage is applied. This voltage is shown in Fig. 38(c). v_1 is in turn applied to the ringing circuit formed by the second L-C section and R . The effect produced is not vastly different from that which would occur if the input pulse were applied to this network, not at time $t = 0$, but at time $t = T$. In other words, the voltage v_0 across R is similar to v_1 except that v_0 does not reach the value of the input voltage until $t = T'$, where T' is approximately equal to $2T$. Subsidiary effects also occur, such as minor alterations in the character of the ringing at the end of the delay time T' . (Fig. 38(d)).

The addition of further L-C sections produces a similar effect. The delay time before the rise in input voltage appears across the output is approximately proportional to the number of stages used. Neither the amplitude of the ringing, nor its duration, is appreciably affected by an increase in the number of L-C sections.

In the simple ringing circuit of Fig. 36(b) the delay T and the period of the ringing are both proportional to \sqrt{LC} . In the multiple-stage network the total delay is proportional to $n\sqrt{LC}$, where n is the number of stages, whilst the period of the ringing is approximately independent of n . By increasing n and reducing L and C proportionately, the ringing can be made negligible without increasing the delay time. If this is done in a six-stage or eight-stage network, the output voltage is made approximately rectangular, as shown in Fig. 38(e).

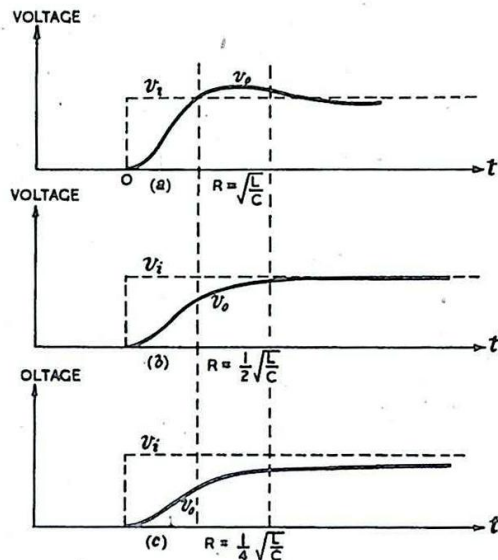
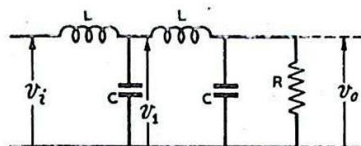
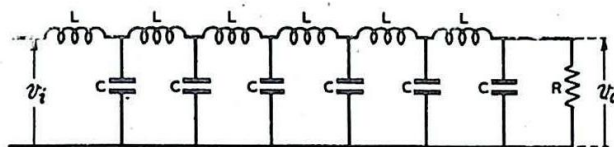


Fig. 37.- Response of L-C-R circuit of fig. 36(b).

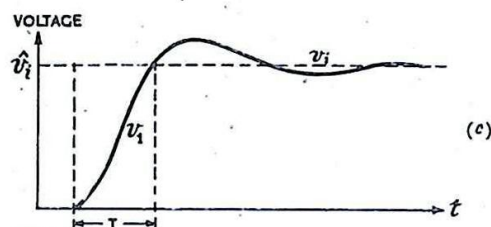


(a)

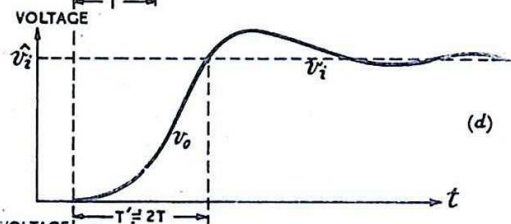


(b)

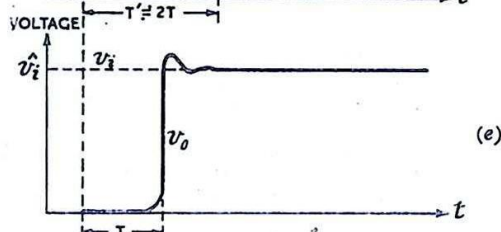
Fig. 38.- Response of delay network.



(c)



(d)



(e)

NON INSTANTANEOUS CHANGES OF VOLTAGE

12. General Discussion

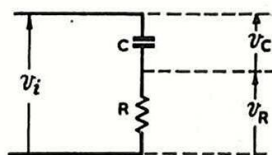
In practice it is impossible to produce an absolutely instantaneous change in voltage, but in many cases no great error is introduced in considering a change of voltage to be instantaneous if the total time taken for the change is of the order of $1/10 \mu\text{s}$. However, sometimes the change of voltage applied across a circuit must be considered as changing linearly with time. A common example is the sawtooth voltage produced by time-base circuits (Fig. 42). Sometimes the rise or fall of voltage is exponential, but usually only a small portion of the exponential rise or fall is considered and for practical purposes this portion may be regarded as linear.

13. Input Voltages Increasing Linearly With Time

Let the input voltage v_i applied to the condenser (capacitance C) and resistor (resistance R) in series increase at a constant and finite rate m , so that

$$v_i = mt$$

The voltage V_R across the resistor rises exponentially and tends to a final value mCR . The initial rate of rise is the same as that of the input voltage v_i .



The voltage v_C across the condenser also begins to rise exponentially, but tends to rise at the same rate as v_i when v_R approaches its steady value. This is indicated in Fig. 39.

Mathematical analysis

At time t let :-

q = charge on condenser

i = current in circuit

$$\left(= \frac{dq}{dt} \right)$$

Then :-

applied voltage = voltage across condenser + voltage across resistor

$$\text{or } v_i = mt = \frac{q}{C} + iR$$

Differentiating with respect to t :-

$$m = \frac{1}{C} \frac{dq}{dt} + R \frac{di}{dt}$$

$$\text{or } m = \frac{i}{C} + R \frac{di}{dt}$$

The solution of this equation is :-

$$i = mC (1 - e^{-t/CR})$$

$$\text{Therefore } v_R = iR = mCR (1 - e^{-t/CR})$$

$$\text{and } v_C = v_i - v_R = mt - mCR(1 - e^{-t/CR})$$

Thus v_R rises exponentially to a limiting value mCR . The smaller is the time-constant CR or the slower the rate of rise of the input voltage (smaller m), the lower does this final value of v_R become. The rate of rise of v_R is

$$\frac{d}{dt} (v_R) = m e^{-t/CR}$$

so that at time $t = 0$ the rate of rise is m , i.e., v_R initially rises at the same rate as v_i . The rate of rise of voltage across the condenser starts from zero and increases until, by the time v_R is constant at a value mCR , it becomes the same as that of the input voltage.

If a voltage which rises linearly with time is applied to a coil and resistor in series, the voltage v_L across the coil and the voltage v_R across the resistor at any instant are given by the expressions

$$v_L = m \frac{L}{R} (1 - e^{-\frac{Rt}{L}})$$

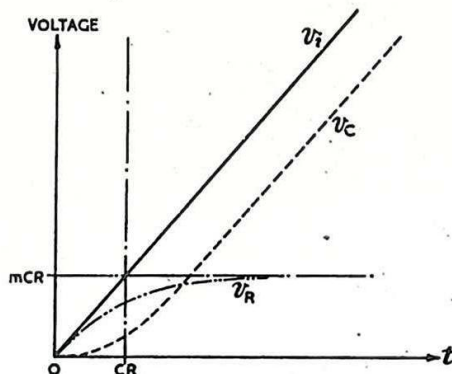


Fig.39. - Response of C-R circuit to voltage varying linearly with time.

$$\text{and } v_R = mt - \frac{mL}{R} (1 - e^{-\frac{Rt}{L}})$$

The voltage across the coil rises towards a final value $\frac{mL}{R}$, and is similar to the voltage across the resistor in the corresponding C-R circuit. The voltage across the resistor in the L-R circuit is similar to that across the condenser in the C-R arrangement.

14. Linear Rise of Voltage of Finite Duration

Now let us suppose that the input voltage applied to a series C-R circuit rises at a rate m for a time T' and then remains constant at a value mT' for an indefinite time; (Fig. 40).

During the time T' the voltage across the resistor rises exponentially towards a maximum value equal to mCR . Throughout this time v_R is given by :-

$$v_R = mCR (1 - e^{-t/CR})$$

During the time the input voltage remains constant v_R falls exponentially from its value at T' towards a final value of zero.

Three cases arise, in which $CR \gtrless T'$. These are illustrated by the following examples corresponding to the three diagrams of Fig. 40. We are concerned here only with the value of v_R during the rise of v_i .

$$(i) \quad CR = 5T'$$

At time T' the value of v_R is $0.9 mT'$, i.e., the voltage across the resistor rises to nine-tenths of the maximum amplitude of the applied voltage. The greater the time-constant, the more closely v_R approaches the amplitude of the input voltage.

$$(ii) \quad CR = T'$$

The value of v_R at the time T' is approximately two-thirds of the maximum amplitude of the input voltage.

$$(iii) \quad CR = \frac{1}{5} T'$$

The value of v_R at the time T' is nearly equal to one-fifth of the maximum amplitude of the input voltage, i.e., the smaller the time-constant the smaller is the value of v_R .

The voltage across the condenser at any instant may be obtained from the relation

$$v_i = v_C + v_R$$

This is illustrated in Fig. 40 .

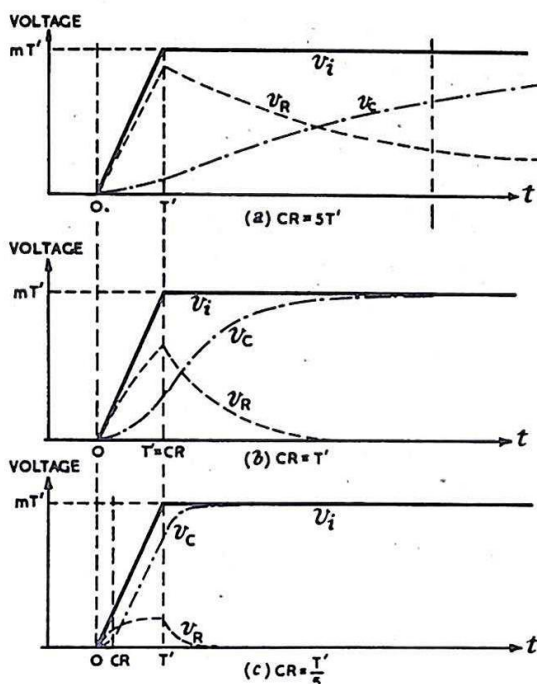


Fig.40. - Response of C-R circuit to linear voltage rise of finite duration.

It should be noted that since there is no instantaneous change in the applied voltage at any time, in particular at $t = 0$ and $t = T'$, there is never any sudden change in the rate of rise of v_C .

15. Pulses Involving Linear Variations of Voltage

Fig. 41 shows the voltage developed across the resistor, and that developed across the condenser, when a trapezium-shaped pulse is applied. A so-called "rectangular" pulse usually has this trapezium shape, i.e., the changes of voltage are not instantaneous. It has been assumed that the input voltage rises to its constant value in an interval T' , remains at this value for a time T , and then falls back to its initial value in a further interval T' . T must be very much greater than T' if the voltage change is to represent the usual type of "rectangular" pulse. When the time-constant of the circuit is smaller than T , short pulses are produced across the resistor (compare Sec. 3). However, if the time-constant is made progressively smaller, the amplitude of these short pulses decreases. When the time-constant is much greater than the time T , the voltage developed across the resistor closely approximates to the input voltage.

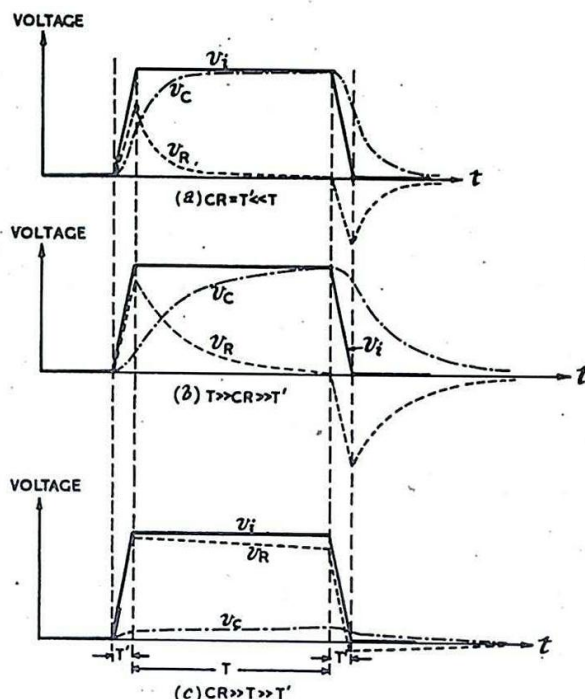


Fig.41.- Response of C-R circuit to trapezoidal pulse.

Fig. 42 shows the voltage developed across the resistor and that developed across the condenser when a pulse of triangular shape is applied. The rise of input voltage takes place in an interval T' , and the fall of input voltage is assumed to be instantaneous. Such a variation is commonly called a Sawtooth Voltage. When the time-constant of the circuit is much greater than T' , the voltage across the resistor closely resembles the input voltage. When the time-constant is much smaller than T' , the voltage variation across the condenser approximates in shape to that of the input voltage, whilst the voltage across the resistor rises to a small value during T' and at $t = T'$ drops sharply to form a negative-going pulse.

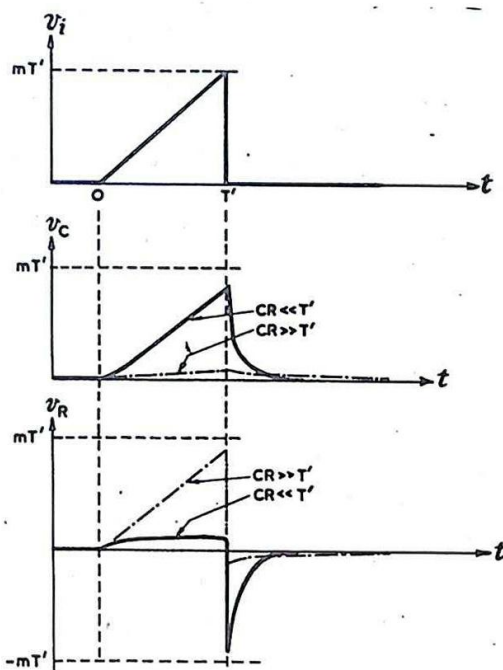


Fig.42.- Response of a C-R circuit to sawtooth pulse.

Fig. 43 shows the voltages developed across the resistor and the condenser when

a triangular pulse which has equal rates of rise and fall of voltage is applied to the circuit. It should be noted that, when the time-constant of the circuit is less than the time of rise or fall of the input voltage, the voltage variation developed across the resistor approximates to a pair of rectangular pulses of small amplitude.

16. Succession of Pulses

As in the case of the rectangular pulses dealt with in Secs. 4 and 8, an input which consists of a succession of uniform pulses of any shape may be divided into a steady component equal to the mean value \bar{v}_i of the input and an alternating component whose mean value is zero. When such an input is applied to a series C-R circuit, an equilibrium condition is ultimately arrived at in which the steady component of the voltage across the condenser is equal to \bar{v}_i whereas the mean value of the voltage across the resistor is zero. The fraction of the alternating component appearing across either the condenser or the resistor, and the amount of distortion present, depend on the time-constant.

Fig. 44 shows the waveforms of some pulses which are commonly applied to a series C-R circuit. Fig. 45 shows the voltages developed in the steady state across the resistor and condenser when a continuous train of triangular pulses is applied to the circuit. Fig. 46 shows the results to be expected when a continuous train of sawtooth pulses is applied.

If a succession of pulses is applied to a series L-R circuit the resulting conditions are comparable with those which obtain in a series C-R circuit. The voltages developed across the coil and resistor are similar to those developed across the resistor and condenser respectively in the C-R arrangement.

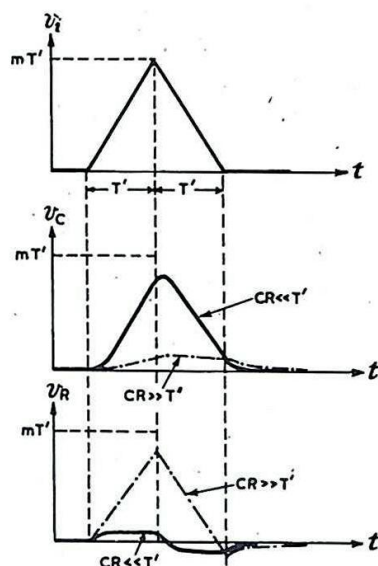


Fig.43.- Response of a C-R circuit to triangular pulse.

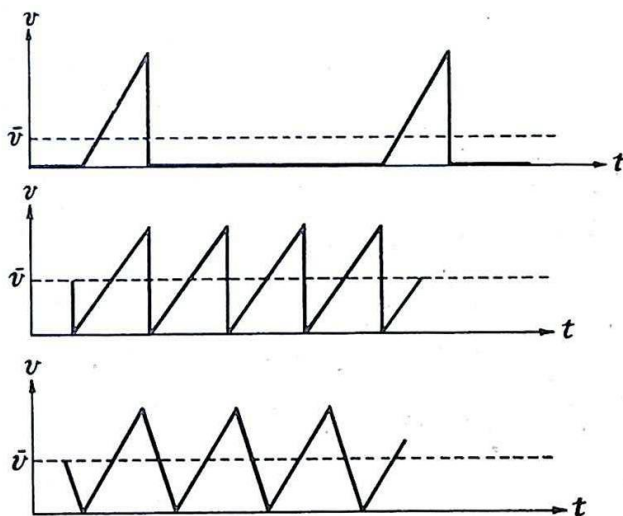


Fig.44.- Typical waveforms

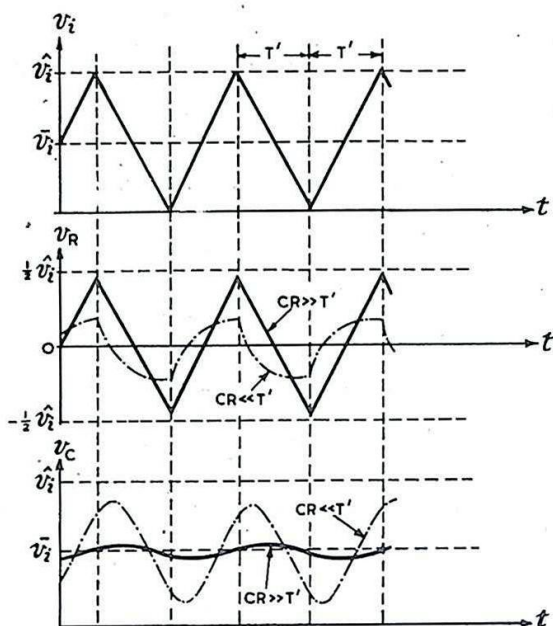


Fig.45.- Response of a C-R circuit to a succession of triangular pulses

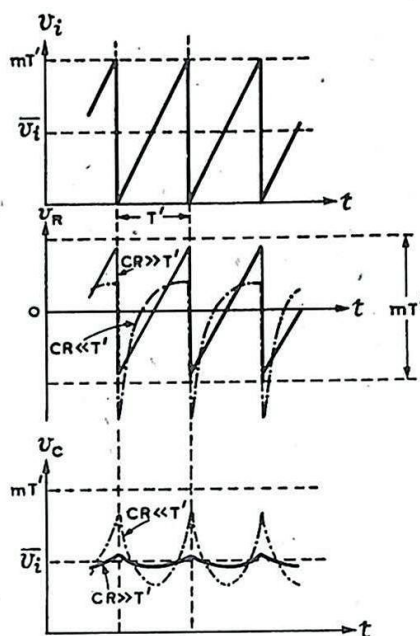


Fig.46.- Response of a C-R circuit to a succession of sawtooth pulses

17. DIFFERENTIATING AND INTEGRATING CIRCUITS

With certain limitations, which are discussed below, it is possible to arrange a circuit so that the output provides either a differential or an integral version of the input voltage.

$$\text{i.e. either } v_o \propto \frac{dv_i}{dt}$$

$$\text{or } v_o \propto \int v_i \cdot dt$$

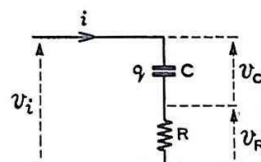
A series C-R circuit may be used to meet either of these requirements. For differentiation the time-constant must be short, and the output is taken across the resistor. For integration the time-constant must be long, and the output is taken across the condenser. The conditions governing the choice of time-constant are investigated below.

Considering the circuit of Fig. 47, we know that

$$v_i = v_C + v_R$$

$$i = \frac{dq}{dt}$$

$$\text{and } v_C = \frac{q}{C}$$



Differentiating Circuit

Assume that $v_R \ll v_C$

$$\text{i.e., } v_C \doteq v_i$$

$$\text{then } q/C \doteq v_i$$

and by differentiation with respect to time, we obtain

$$\frac{1}{C} \frac{dq}{dt} \doteq \frac{d}{dt} (v_i)$$

Fig.47.- A differentiating or integrating circuit.

$$\text{Hence } \frac{i}{C} \doteq \frac{d}{dt} (v_i)$$

$$\text{or } \frac{v_R}{CR} \doteq \frac{d}{dt} (v_i)$$

$$\text{i.e. } v_R \doteq CR \frac{d}{dt} (v_i)$$

In words, the voltage across the resistor is approximately proportional to the time-rate-of-change of input voltage.

We shall now consider the circumstances under which the assumption

$$v_R \ll v_C$$

is justified.

$$v_R = Ri = \frac{Rdq}{dt} = CR \frac{d}{dt} (v_C).$$

Hence the assumption may be written

$$CR \frac{dv_C}{dt} \ll v_C$$

But, as shown

$$v_C \doteq v_i ;$$

hence

$$CR \frac{dv_i}{dt} \ll v_i ,$$

i.e.

$$CR \ll v_i / \frac{dv_i}{dt}$$

This means that the time-constant must be small compared with the ratio of magnitude to time-rate-of-change of the input voltage.

Integrating Circuit

$$\text{Assume that } v_C \ll v_R ,$$

$$\text{i.e. } v_R \doteq v_i ;$$

$$\text{then } Ri \doteq v_i$$

and by integration with respect to time we obtain

$$R \int i dt \doteq \int v_i \cdot dt$$

$$\text{Hence } Rq \doteq \int v_i \cdot dt$$

$$\text{or } CR v_C \doteq \int v_i \cdot dt$$

$$\text{i.e. } v_C \doteq \frac{1}{CR} \int v_i \cdot dt$$

In words, the voltage across the condenser is approximately proportional to the time-integral of the input voltage.

It may be shown, by reasoning similar to that used in the case of the differentiating circuit, that the assumption

$$v_C \ll v_R$$

implies that

$$CR \gg \frac{\int v_i dt}{v_i}$$

This means that the time-constant must be large compared with the ratio of the integrated value to the magnitude of the input voltage.

As an example of differentiating we may take the circuit and waveform of Fig. 39

Here $\frac{dv_i}{dt} = m$ and is constant, so that v_i increases steadily with time. Hence if t is large enough

$$v_i / \frac{dv_i}{dt} \gg CR$$

and the output v_R is a differentiated version of v_i . This is shown in the figure, where, for values of $t \gg CR$, $v_R \propto m$

As an example of integration, consider the waveforms of Fig. 29 For values of $t \ll CR$ the condenser voltage continues to rise with each pulse, so that

$$v_C \propto \int v_i dt$$

For larger values of t this condition no longer holds and eventually a steady state is reached. Under these circumstances

$$\int v_i dt \gg CR v_i$$

and the condition for integration is not fulfilled.

The terms "differentiating" and "integrating" are often applied to short and long time-constant circuits respectively, even though they are not used in the conditions under which true differentiation or integration occurs. For example, the voltage v_R shown in Fig. 48 is frequently referred to as a differentiated version of the input v_i . In fact $\frac{dv_i}{dt}$ is either infinite or zero, as shown. Provided $CR \ll T$, v_R consists of narrow peaks of voltage and the similarity with the true differentiated voltage is obvious.

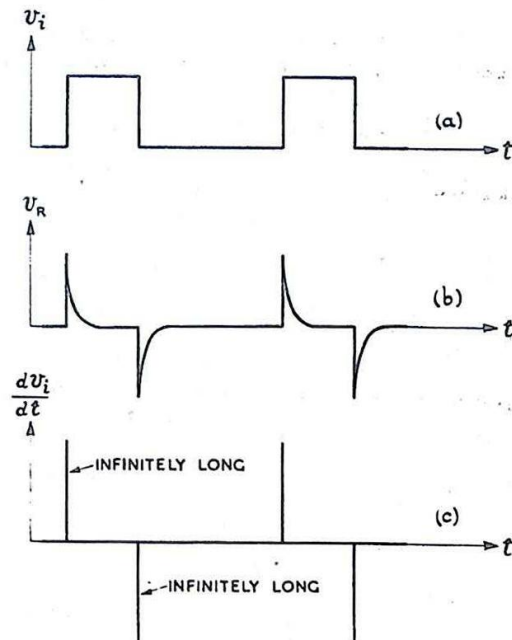


Fig.48.- Comparison between output (b) of a "differentiating" circuit and a true differentiated version (c) of the input (a).

CHAPTER 3

NETWORKS

1. INTRODUCTION

In all radar circuits involving the transmission of electrical energy, examination of the circuit behaviour at any stage requires the division of the system into "source" and "load", as shown in Fig. 49. If the system is linear it is conveniently described in terms of ideal

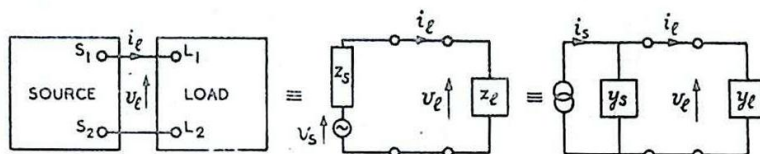


Fig.49.- Division of system into source and load.

generators and input and output impedances, as shown in Chap. 1, Sec. 12. The voltage v_l developed across the load, the current i_l through it, and the phase relation between them specify the load or input impedance z_l to a sinusoidal EMF of given frequency. The source may be represented by a generator of EMF equal to the voltage v_s between the output terminals $S_1 S_2$ when open-circuited, in series with the equivalent source impedance z_s , called its output impedance. Alternatively the source may be

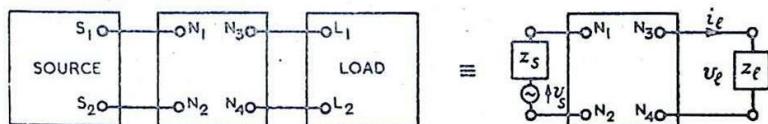


Fig.50.- Network interposed between source and load.

represented by a generator of current i_s equal to the current through the short-circuited output terminals, in parallel with its output impedance z_s (or output admittance $y_s = \frac{1}{z_s}$)

An element in the chain between source and load may usually be represented as a four-terminal network (Fig. 50). Such a network may contain linear or non-linear elements, e.g. transmission lines, valve amplifiers, etc. For simplicity, only those networks will be considered in any detail which do not contain sources of EMF and which are formed of resistive (R), inductive (L) and capacitive (C) elements. For one-way transmission of energy, from source to load, such a network may be represented by a potentiometer

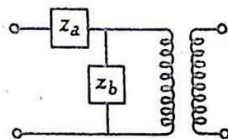


Fig.51.- Representation of 4-terminal network for one-way transmission.

followed by an ideal 1:1 transformer, (Fig. 51); whilst if two-way transmission is to be considered the four-terminal network may be represented by either a T-section or a π -section followed by an ideal 1:1 transformer, as shown in Fig. 52.



Fig. 52. - Representation of a 4-terminal network.

The input and output conditions so far described have ignored any connections between source and load except through the network. In practice, the relations between the potentials at the input terminals and of neighbouring conductors ("earth") are important, and must correspond to the relations between the potentials at the output terminals and "earth". In general, this is equivalent to adding extra terminals to the network, as shown in Fig. 53(a). Two cases are commonly encountered :-

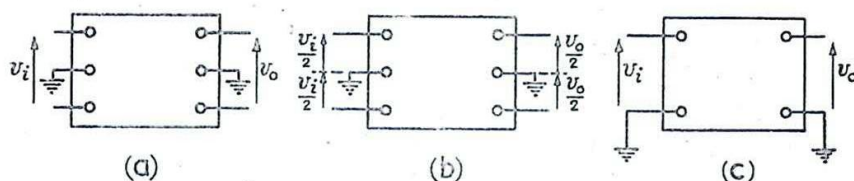


Fig. 53. - Effect of earth: balanced and unbalanced networks.

- (i) **Balanced networks:** in these, the potentials at the input and output terminals are symmetrically disposed with reference to earth (Fig. 53(b))
- (ii) **Unbalanced networks:** in these, one each of the input and output terminals is earthed (Fig. 53(c)).

The properties of balanced networks can readily be derived from those of unbalanced networks, and only the latter will be considered in subsequent paragraphs. For example, the fundamental properties (ie. Characteristic Impedance, Propagation Constant, etc.) of the balanced H-network of Fig. 54 are the same as those of the T-network of Fig. 55(a) provided that the total shunt impedance z_2 and the total series impedance z_1 are the same for the two networks, as indicated.

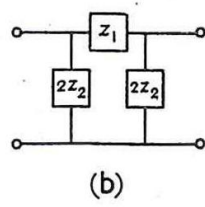
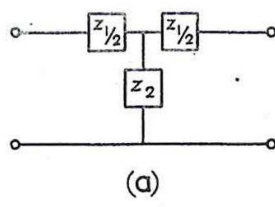
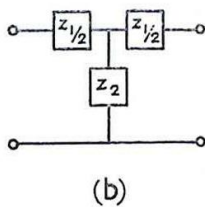
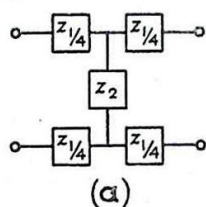


Fig. 54. - Balanced and unbalanced networks.

Fig. 55. - Symmetrical T- and π -networks.

Analysis and design are very much simplified if the network can be made symmetrical. This means that if the network is reversed, and its "output terminals" connected to the source and its "input terminals" connected to the load its properties remain the same, whatever the nature of the source or load. The advantage is particularly marked in the case of iterative networks of the types shown in Fig. 56. If many sections are used, with a particular termination, the behaviour of the network as a whole can be described in terms of the change of phase and attenuation due to each section, irrespective of whether the network is split up into T- or π -sections, symmetrical or unsymmetrical. The type that is most simply described, namely the symmetrical section, is thus the most useful to consider and is of sufficiently general application. An unsymmetrical network can always be considered as a symmetrical one with an extra series or shunt component added at the input or output terminals, or both. Symmetrical T- and π -sections are illustrated in Fig. 55.

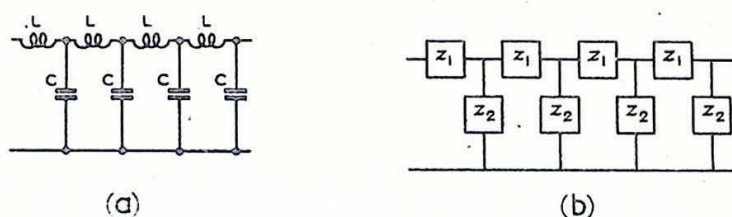


Fig. 56. - Iterative networks.

Some of the properties of unbalanced symmetrical networks will be considered in subsequent sections; only the simplest types will be dealt with in any detail.

2. MEASUREMENT OF INPUT AND OUTPUT IMPEDANCES

In the circuit of Fig. 49(a) the input impedance of the load may be determined by direct measurement of the magnitudes of v_l and i_l and their phase relationship. Theoretically, the output impedance of the source is determined by measuring the input impedance between S_1 and S_2 with the generator suppressed (i.e., prevented from generating but without change of its internal impedance.) The remainder of the circuit must be unaltered. From a practical point of view this experiment is one which cannot be performed in the majority of cases: if, through any cause, the actual generator ceases to generate, its internal impedance is almost inevitably altered by this change in the physical conditions. A method which is capable of practical application will now be described.

Suppose that the output impedance of the source is capacitive. A variable load is connected as shown in Fig. 57, X_l in this case being the reactance of a variable inductance. As the load inductance is varied the magnitude of the alternating voltage v_R developed across the load

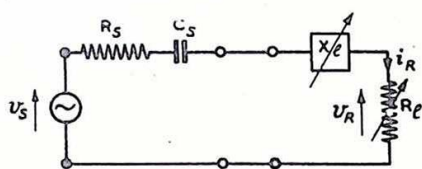


Fig. 57. - Circuit for measuring generator voltage and impedance.

resistance will vary, and will reach a maximum value when the inductive reactance $X_L = \omega L$ of the load exactly nullifies the capacitive reactance $- \frac{1}{\omega C_S}$ of the source. A similar method is applicable if the output impedance is inductive.

The resistive component of the output impedance may then be determined by varying the load resistance, and comparing the voltage v_R developed across it with the current i_R through it. This variation is shown in Fig. 58. The L-C combination may be considered as a short-circuit as far as v_R and i_R are concerned. When sufficient points on the graph have been obtained, \hat{v}_S and \hat{i}_S , and hence R_S , may be determined from the intercepts on the axes, since $R_S = \frac{\hat{v}_S}{\hat{i}_S}$.

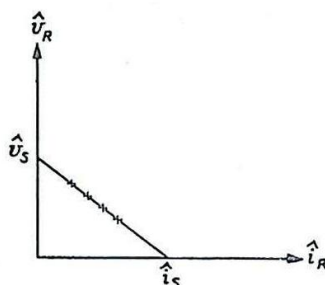


Fig.58.- Determination of output resistance and open-circuit voltage.

Alternatively, the method shown in Fig. 59 may be employed, a similar graph being derived (Fig. 60). It should be noted that i_G , v_G , G_L , B_L , are not the same as i_R , v_R , $\frac{1}{R_L}$, $\frac{1}{X_L}$ (See Chap.1, Sec.17)

Practical difficulties arise because, where the generator is an inherent part of the source, changes in load affect its frequency as well as its output voltage and impedance. To reduce these effects to negligible magnitudes, the changes in load impedance should be made very small compared with the total impedance of the circuit.

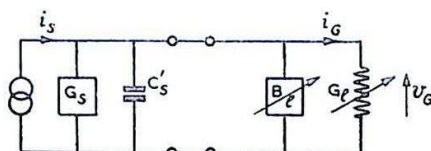


Fig.59.- Circuit for measuring generator current and admittance.

3. SYMMETRICAL T-SECTIONS AND II-SECTIONS

Whatever the actual arrangements of the elements forming an unbalanced four-terminal network, it is convenient to describe the behaviour of the network in terms of the equivalent T- or II-network of Fig. 55.

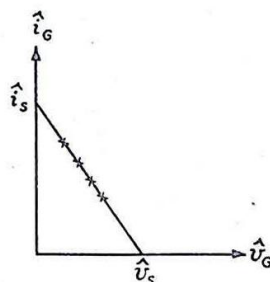


Fig.60.- Determination of output conductance and short-circuit current.

The input impedance z_s of a symmetrical T-network depends on the manner of its termination z_r according to the formula :-

$$z_s = \frac{z_1}{2} + \frac{z_2 \left(\frac{z_1}{2} + z_r \right)}{z_2 + \frac{z_1}{2} + z_r}$$

(Fig. 61(a))

It is not always desirable to work in terms of z_1 and z_2 , and three other impedances are frequently used:-

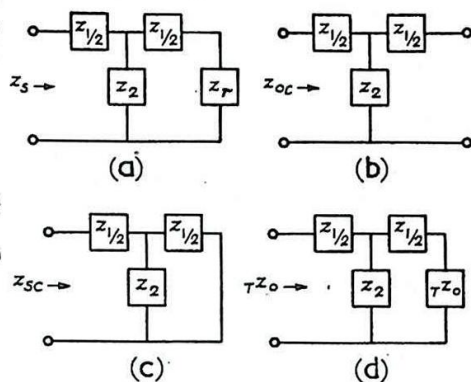


Fig. 61.- Input impedance of T-network for various standard terminations.

(i) Open-circuit impedance:

z_{oc} ; with the output terminals open-circuited,

$$rZ_{oc} = \frac{z_1}{2} + z_2 \quad (\text{Fig. 61(b)})$$

(ii) Short-circuit impedance :

z_{sc} ; with the output terminals short-circuited,

$$rZ_{sc} = \frac{z_1}{2} + \frac{z_2 \cdot \frac{z_1}{2}}{z_2 + \frac{z_1}{2}} \quad (\text{Fig. 61(c)})$$

(iii) Characteristic impedance:

rZ_o ; this is a particular impedance such that when the network is terminated in rZ_o its input impedance is rZ_o .

$$rZ_o = \frac{z_1}{2} + \frac{z_2 \left(\frac{z_1}{2} + rZ_o \right)}{z_2 + \frac{z_1}{2} + rZ_o} \quad (\text{Fig. 61(d)})$$

This equation for rZ_o reduces to $rZ_o = \pm \sqrt{z_1 z_2 + \frac{z_1^2}{4}}$

The ambiguity of the \pm sign is important in the case of reactive impedances, and must be resolved by further consideration of the particular circuit.

π -network

Corresponding expressions for the π -network of Fig. 62 are :-

(i) Open-circuit admittance

$$y_{oc} = \frac{y_2}{2} + \frac{\frac{y_2}{2} \cdot y_1}{\frac{y_2}{2} + y_1}$$

(ii) Short-circuit admittance

$$y_{sc} = \frac{y_2}{2} + y_1$$

(iii) Characteristic admittance ...

$$\pi y_o = \sqrt{y_1 y_2 + \frac{y_2^2}{4}}$$

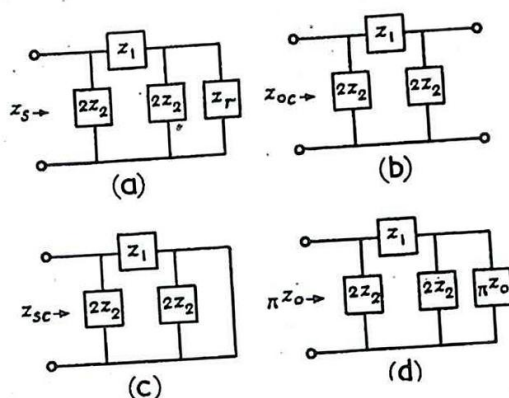


Fig. 62.- Input impedances of π -network for various standard terminations.

General

Results frequently useful are :-

$z_o = \sqrt{z_{sc} z_{oc}}$, both for T- and π -networks; and
 $T z_o \cdot \pi z_o = z_1 \cdot z_2$, where z_1 and z_2 are the same for both networks.

Propagation constant

When the network is terminated in its characteristic impedance, the ratio between input and output voltages and currents is the same. This ratio is expressed

$$\frac{i_s}{i_r} = \frac{v_s}{v_r} = \epsilon^{\gamma} \quad \text{where } \gamma \text{ is called the Propagation Constant of the network.}$$

In general, γ possesses real and imaginary parts, and is equal to $\alpha + j\beta$, where α corresponds to the attenuation, and β the phase-shift introduced by the properly terminated section.

The magnitude of ϵ^{γ} will be denoted by k , so that $\alpha = \log_e k$.

It may be shown that for any symmetrical T- or π -network, γ satisfies the relation

$$\epsilon^{\gamma} + \epsilon^{-\gamma} = 2 + \frac{z_1}{z_2} \quad \text{or}$$

4. MATCHING

Consider a transmission system divided into source and load as in Fig. 49. For a given source the load may be designed to satisfy one of many requirements. The three principal requirements are :-

- (i) Maximum current supplied to the load;
- (ii) Maximum voltage applied to the load;
- (iii) Maximum power transferred to the load.

In transmission lines and waveguides a further requirement is the avoidance of reflections and standing waves; this will be dealt with in the appropriate chapters. The load impedances which satisfy the three principal requirements for linear networks are, respectively :-

$$(i) R_L = 0, X_L = -X_S;$$

$$(ii) G_L = 0, B_L = -B_S; \text{ (i.e. } R_L = 0, X_L = -\frac{R_S^2 + X_S^2}{X_S} \text{)}$$

$$(iii) R_L = R_S, X_L = -X_S; \text{ or alternatively, } G_L = G_S, B_L = -B_S.$$

When the conditions (iii) are satisfied, the load and source impedances are said to be conjugate, and the load is said to be matched to the source, and vice versa.

For non-linear networks other criteria must be employed. The power delivered to the load, for instance, may be plotted against the load resistance, and a series of curves obtained for different values of load reactance. The optimum values of R_L and X_L are obtained as shown in Fig. 63. Where the regulation of the source with changes in loading is an important consideration, it may be desirable not to choose the load input impedance for maximum power, but to take the stability into account (E.g. Chap. 8 Sec. 40).

If source and load are not variable, and the ideal matching conditions not realised, a matching section may be inserted in the line between source and load. Ideally, this will consist of reactive elements, designed to provide maximum power transfer from source to load without itself absorbing energy. (Since the input impedance of the load is fixed, the conditions for maximum power also ensure that both maximum voltage and current are delivered to the load). The output impedance of the network, fed by the source, will be the conjugate of z_L , whilst the input impedance of the network, terminated in z_L , will be the conjugate of z_S , as indicated in Fig. 64. In general it will not be possible to satisfy these conditions exactly except at one or two frequencies, although approximate matching may be obtained over a band of frequencies.

For resistive impedances at audio frequencies an iron-cored transformer is a suitable matching device. This has the advantage that its matching properties for low frequencies are reasonably independent of frequency, and depend only on the ratio of R_S to R_L . For use at higher frequencies various types of matching sections are used, all suffering more or less from the disadvantage of being sensitive to frequency changes. Some of these are described in Chaps. 4 and 5.

5. GAIN AND LOSS

In general the network of Fig. 50 will involve either amplification or attenuation of the input voltage, current and power. If the

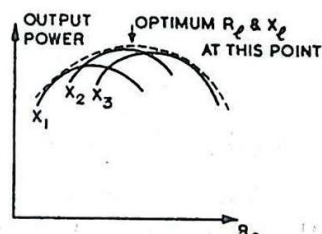


Fig. 63.- Method of determining correct matching conditions when system is not linear.

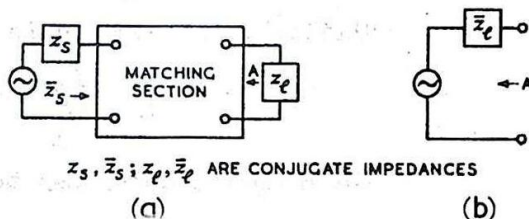


Fig. 64.- Use of matching section.

magnitude of the sinusoidal input voltage is \hat{v}_i and of the output voltage \hat{v}_o , then \hat{v}_o / \hat{v}_i is the voltage amplification of the network. Similarly $\frac{\hat{i}_o}{\hat{i}_i}$ and $\frac{P_o}{P_i}$ are the current and power amplifications.

It is customary to define quantities known as "Gains and Losses" in terms of the logarithms of these ratios.

Power ratio The Power Gain is defined as

$$g_p = \log_{10} \frac{P_o}{P_i} \text{ bels}$$

$$= 10 \log_{10} \frac{P_o}{P_i} \text{ decibels (db.)}$$

Voltage ratio

The Voltage Gain is defined as

$$g_v = \log_{\epsilon} \frac{\hat{v}_o}{\hat{v}_i} \text{ nepers}$$

Current ratio The Current Gain is defined as $g_i = \log_{\epsilon} \hat{i}_o / \hat{i}_i$ nepers.

Other terms are also used. Thus, voltage and current gains are commonly quoted in decibels. The appropriate definitions are :-

$$g_v = 20 \log_{10} \frac{\hat{v}_o}{\hat{v}_i} \text{ decibels}$$

$$g_i = 20 \log_{10} \frac{\hat{i}_o}{\hat{i}_i} \text{ decibels}$$

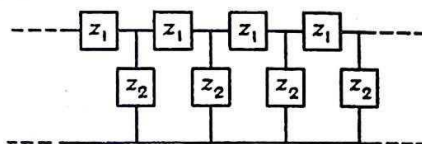


Fig.65.- Iterative networks.

These definitions imply that if the input and load impedances are identical the voltage, current and power gains give the same numerical value in decibels.

Similarly the power gain may be quoted in nepers, in which case

$$g_p = \frac{1}{2} \log_{\epsilon} \frac{P_o}{P_i} \text{ nepers}$$

These equivalences can be summarised by the statements

$$1 \text{ neper} = 8.69 \text{ decibels}$$

$$\text{or} \quad 1 \text{ decibel} = 0.115 \text{ nepers.}$$

Where attenuation instead of amplification occurs the corresponding formula for the decibel or neper loss may be obtained from the above by interchanging the subscripts i and o.

e.g. the voltage loss is $20 \log_{10} \frac{\hat{v}_i}{\hat{v}_o}$ decibels

6. ITERATIVE SECTIONS

Four-terminal networks are frequently used in Cascade (or Tandem) to form a line. Such a line is shown in Fig. 65 and may be called an Artificial Line, as distinct from a uniform transmission line which it

may be desired to simulate. It may be considered as an iterative sequence of either T- or Π -sections, according to the mode of termination, as illustrated in Fig. 66. In the mechanical construction of the complete network no distinction is made between T- and Π -sections except at the output and input terminals. Since it is usual to match the input to the source by a matching section, it is convenient to consider the output termination only as determining the manner in which the line is divided into its constituent networks for purposes of analysis. Thus the network of Fig. 67(a) may be split into two parts, as in Fig. 67(b), the iterative portion terminated in its characteristic impedance rZ_0 , and the non-iterative portion at the input, each part being considered separately.

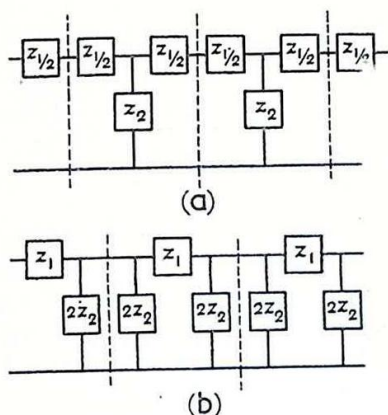


Fig. 66.- Dividing an iterative network into symmetrical T- and Π -sections.

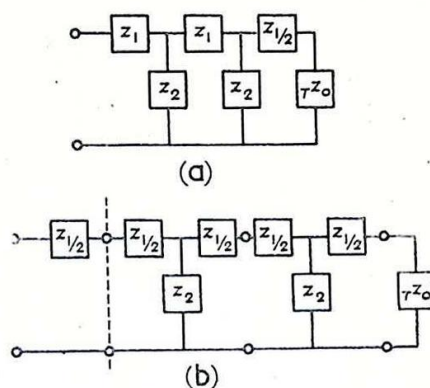


Fig. 67.- Dissection of network for purposes of analysis.

If a line is terminated in its characteristic impedance appropriate to the mode of division, it exhibits the input properties of an infinite line, since any number of sections may be inserted between the output terminals and the load without interfering with the input conditions. It follows that although the characteristic impedance depends on the mode of termination and division, the propagation constant of a single section is the same provided the line is properly terminated. For example, in the network of Fig. 68 the attenuation and phase shift per section are characteristic of the infinite line (b) and are therefore the same for the iterative T-network of (a) as for the equivalent Π -network of (d).

(The Sign \equiv used in this and other diagrams signifies that corresponding voltage and currents in the various networks are the same)

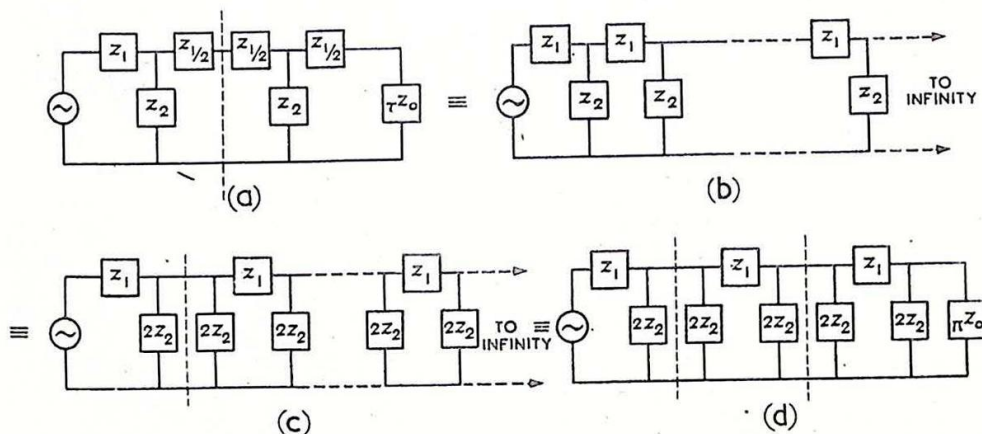


Fig. 68.- Alternative representations of a properly matched iterative network.

Iterative networks are commonly used as attenuators, filters or delay networks, and the particular properties of these will be described in subsequent sections.

7. ATTENUATORS

An attenuator is required to yield at the output terminals an undistorted fraction of the input power or voltage.

(For an ideal attenuator γ , the propagation constant, is real, so that $\beta = 0$ (See Sec. 3); also the characteristic impedance is independent of frequency). To meet the requirements with the symmetrical networks described above, z_1 and z_2 must be the same kind of impedance (i.e. the phase angles must be the same), and should be chosen to suit the load. Since the load is normally resistive, z_1 and z_2 are usually resistances, and, particularly when the attenuator is used for high frequency measurements, special care must be taken to avoid stray reactance. Under certain circumstances, such as for attenuating a large alternating voltage to apply to the deflector plates of a CRT a capacitive attenuator may be used; whilst for monitoring RF waveforms without extracting power, inductive attenuators may be preferable. Only the resistive type will be considered here; the formulae for these are given below. The corresponding formulae for the cases where z_1 and z_2 are pure reactances of like kind may be obtained by substituting X for R throughout.

Properties of a symmetrical resistive attenuator network

For a network of given type, determined by R_1 and R_2 and divided into symmetrical sections as shown in Fig. 69,

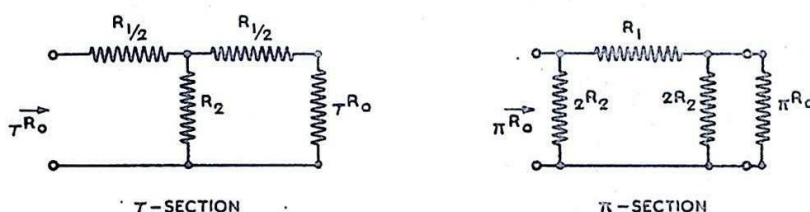


Fig.69.- Matched attenuators.

$$T R_o = \sqrt{R_1 R_2 + \frac{R_1^2}{4}}, \quad G_o = \sqrt{G_1 G_2 + \frac{G_2^2}{4}}$$

$$\text{(Note: } G_o = \frac{1}{R_o}, G_1 = \frac{1}{R_1}$$

$$\text{and } G_2 = \frac{1}{R_2} \text{)}$$

$$\text{and } T R_o \cdot \pi R_o = R_1 \cdot R_2$$

Putting $\beta = 0$ in the formula for γ given in Sec. 3, and writing $\xi^r = \xi^a = k$, we have

$$k + \frac{1}{k} = 2 + \frac{R_1}{R_2}$$

The following useful formulae may be deduced :-

$$k = \frac{R_1 + 2T R_0}{R_1 - 2T R_0} = \frac{\pi R_0 + 2 R_2}{\pi R_0 - 2 R_2} ;$$

$$R_1 = 2 \cdot \frac{k-1}{k+1} \cdot T R_0 \quad \text{and} \quad R_2 = \frac{2k}{k^2-1} \cdot T R_0 ;$$

$$R_1 = \frac{k^2-1}{2k} \pi R_0 \quad \text{and} \quad R_2 = \frac{1}{2} \cdot \frac{k+1}{k-1} \cdot \pi R_0.$$

These formulae enable $T R_0$ or πR_0 and k to be determined from R_1 and R_2 , or vice versa.

The attenuation in db per section is $20 \log_{10} k$.

Examples

- (1) Construct a T-section attenuator to provide 14 db loss per section and to match to a 2000 ohms line.

We have $20 \log_{10} k = 14$.

Hence $k = 5$.

Also $T R_0 = 2000$.

Hence $R_1 = 2 \left(\frac{k-1}{k+1} \right) \cdot T R_0$

$$= 2 \cdot \frac{4}{6} \cdot 2000$$

$$= 2670 \text{ ohms.}$$

Also $R_2 = \frac{2k}{k^2-1} \cdot T R_0$

$$= \frac{10}{24} \cdot 2000$$

$$= 830 \text{ ohms.}$$

Hence each section is formed of two series resistances, each 1335 ohms and a shunt resistance 830 ohms.

- (2) A π -section attenuator has shunt resistances 4000 ohms and series resistance 1000 ohms. To find the characteristic resistance and db loss per section.

We have $R_1 = 1 \text{ k } \Omega$. $R_2 = 2 \text{ k } \Omega$.

Hence $\pi G_0 = \sqrt{1 \cdot \frac{1}{2} + \left(\frac{1}{2}\right)^2 \cdot \frac{1}{4}} \text{ millimhos.}$

$$= \sqrt{\frac{9}{16}} \text{ millimhos.}$$

$$\text{i.e. } R_0 = \frac{4}{3} \text{ k ohms}$$

$$\underline{\underline{= 1330 \text{ ohms.}}}$$

$$\begin{aligned} \text{Also } k + \frac{1}{k} &= 2 + \frac{R_1}{R_2} \\ &= 2 + \frac{1}{2} . \end{aligned}$$

Hence $k = 2$ and the db. loss per section is 6.

FILTERS

8. Ideal Requirements

The ideal requirements for a filter network are illustrated in Fig. 70. The frequency spectrum is divided into pass bands and attenuation bands. In the former, no attenuation is desired, and any unavoidable phase shift should be proportional to frequency. Outside the pass bands there should be infinite attenuation of signals at all frequencies.

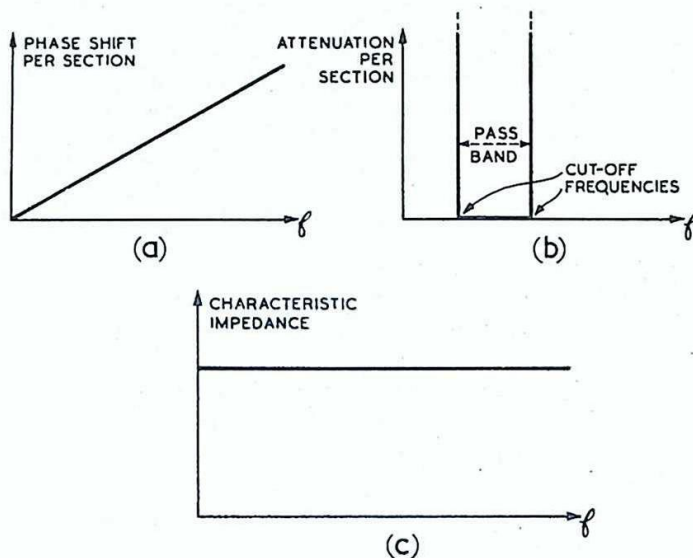


Fig.70.- Ideal characteristics for a filter.

Since the load is normally resistive, the characteristic impedance in the pass band should be a resistance of constant value.

In practice it is impossible to make the phase shift proportional to frequency over the whole of the pass band, and the transition from pass band to attenuation band at the Cut-Off Frequency, as the bordering frequency is called, involves a more or less gradual increase in attenuation as the frequency recedes from its cut-off value.

9. Simple High-Pass and Low-Pass Filters

In the ideal networks about to be described it is assumed that reactive elements are devoid of resistance. The results obtained in

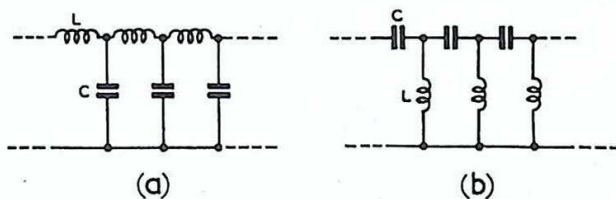


Fig.71.- Simple filter networks.

practice agree closely with those derived from this assumption provided high-Q components are used.

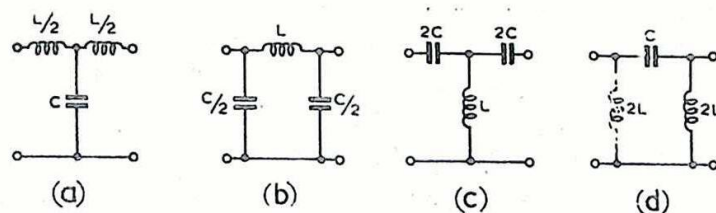


Fig. 72. - Symmetrical T-section and π -section filter networks.

To provide frequency discrimination without attenuation, both capacitive and inductive reactances must be used. The simplest, or prototype, filters are shown in Fig. 71, and the symmetrical T- and π -section arrangements in Fig. 72. For low-frequency filters, the resistance-capacitance networks of Fig. 73 may be used. In these the usual object is to make the capacitive impedances sufficiently low compared with the resistance, at all frequencies likely to be present, that the filter may be considered as passing without attenuation all alternating voltages, but "blocking" steady supplies.

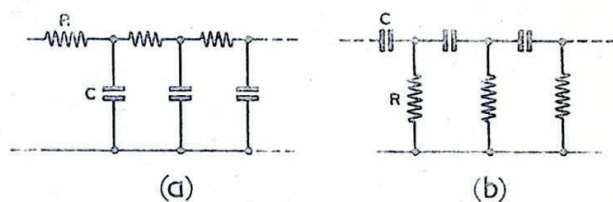


Fig. 73. - R-C filters.

10. Simple Low-Pass Filter Formed of Purely Reactive Elements

Fig. 74(a) represents an iterative network of any number of identical sections formed of perfect coils and condensers; the network is properly terminated so that, at the input terminals PO, it looks like an infinite line.

The letters Q, S, U, denote the spaces of the networks, and the currents, i_Q , i_S etc. the corresponding series currents. v_Q and v_S etc.

are the voltages developed across the series impedances.

The subscript m denotes the electrical centre of a series reactance (e.g. Q_m is the electrical centre of PR) or the space which divides a shunt susceptance into two equal

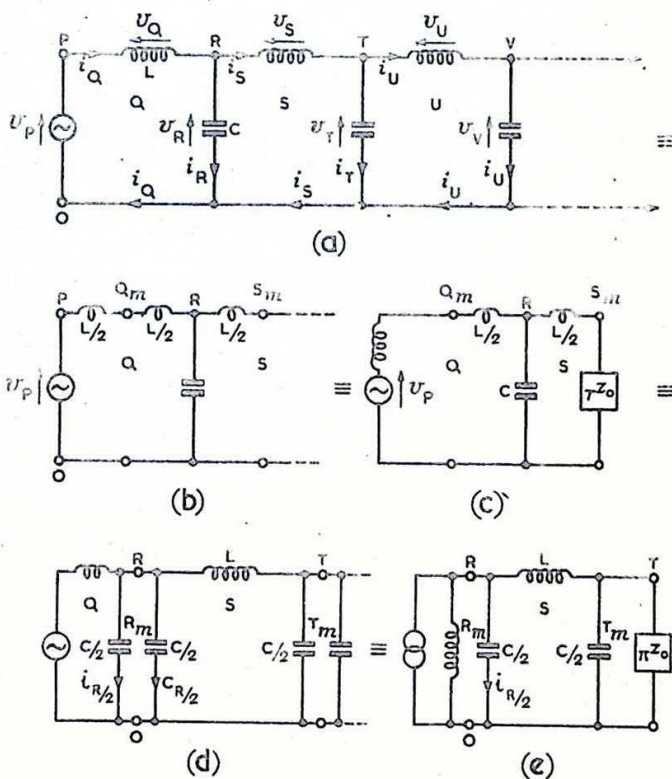


Fig. 74. - Alternative representations of a properly matched iterative network.

susceptances (e.g. R_m divides ωC into two susceptances each $\frac{\omega C}{2}$)

The sign \equiv implies that those portions of the various figures which are similarly lettered carry identical currents and voltages.

In the associated diagrams of Figs. 75 - 77 OP, OR etc. denote the voltages between the corresponding points of Fig. 74, whilst OQ, OS etc. denote the currents in the corresponding branches.

As shown in Section 6, the phase-shift, β radians, and attenuation constant k , for each section, are independent of the manner in which the network is divided into T- or π - sections. To consider the properties of the symmetrical T- network of this type of filter, the line is divided at the middle of one of the series inductances, as shown in Fig. 74(b), where the inductance L between P and R is bisected at Q_m . For a symmetrical π -section each of the shunt capacitances is split into two equal parts as shown in Fig. 74(d), so that the line appears as in Fig. 74(e). For either method of division there is no change in any of the voltages or currents to the right of the chosen input terminals.

The properties of the low-pass filter will be derived by considering the geometry of the vector diagrams. The data upon which these diagrams are based are :-

- (i) The phase shift β radians per section;

i.e., v_S lags v_Q , v_T lags v_R , v_U lags v_S etc., all by β radians.

- (ii) The attenuation k per section; i.e.,

$$k = \frac{\hat{v}_P}{\hat{v}_R} = \frac{\hat{v}_R}{\hat{v}_T} = \dots = \frac{\hat{v}_Q}{\hat{v}_S} = \frac{\hat{v}_S}{\hat{v}_U} = \dots = \frac{\hat{i}_Q}{\hat{i}_S} = \frac{\hat{i}_S}{\hat{i}_U} \dots \text{etc}$$

where \hat{v}_P , etc., are the amplitudes of the alternating voltages and currents.

- (iii) The voltage vector leads the current vector by $\frac{\pi}{2}$ radians for an inductance, and lags by $\frac{\pi}{2}$ radians for a capacitance.

There are two types of vector diagram consistent with these relationships, relating to the pass band and attenuation band respectively.

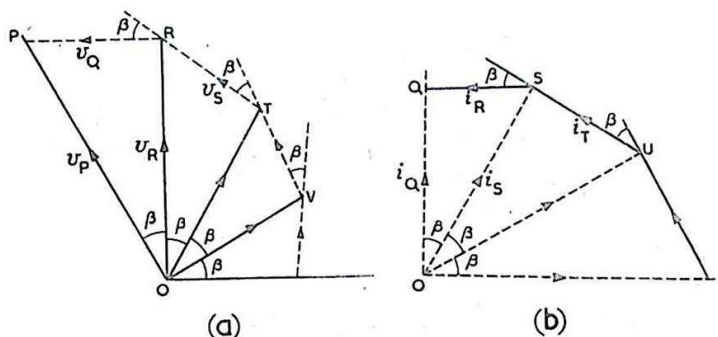


Fig. 75.- Vector diagram.

Pass Band

The relationships (i) and (iii) above are illustrated in the vector diagrams of Fig. 75. It will now be shown that (ii) implies that

$$OP = OR = OT, \text{ etc and } OQ = OS = OU, \text{ etc.}$$

Since FR is perpendicular to OQ (voltage and current vectors for coil FR) and RO is perpendicular to QS (voltage and current vectors for condenser RO).

$$\angle PRO = \angle SQO.$$

Hence the triangles PRO, SQO and, similarly, all the other triangles, are equiangular, so that their corresponding sides are proportional.

Therefore

$$\text{i.e. } \frac{\frac{PO}{RO}}{\frac{\hat{V}_P}{\hat{V}_R}} = \frac{SO}{QO} \dots\dots\dots (1).$$

$$\text{But } \frac{\hat{V}_P}{\hat{V}_R} = \frac{\hat{I}_Q}{\hat{I}_S} = \frac{QO}{SO} = k \text{ from (ii)}$$

$$\text{hence } \frac{QO}{SO} = \frac{SO}{QO}; \text{ i.e. } SO = QO \text{ and } k = 1$$

hence there is no attenuation in the Pass Band.

Phase-Shift

Fig. 76 shows the vector diagrams superimposed, and with these relationships satisfied. Referring to this figure

$$\sin \frac{\beta}{2} = \frac{\frac{1}{2} \frac{\hat{I}_R}{\hat{I}_Q}}{\frac{\hat{V}_Q}{\hat{V}_R}} = \frac{\frac{1}{2} \frac{\hat{V}_Q}{\hat{V}_R}}{\frac{\hat{I}_R}{\hat{I}_Q}} = \sqrt{\frac{\frac{\hat{I}_R}{\hat{I}_Q} \frac{\hat{V}_Q}{\hat{V}_R}}{4}};$$

$$\text{but } \frac{\hat{V}_Q}{\hat{I}_Q} = \omega L, \text{ and } \frac{\hat{V}_R}{\hat{I}_R} = \frac{1}{\omega C};$$

$$\text{hence } \sin \frac{\beta}{2} = \sqrt{\frac{\omega^2 LC}{4}} = \sqrt{\frac{\omega^2}{\omega_c^2}} = \frac{f}{f_c},$$

$$\text{where } \omega = 2\pi f \text{ and } \omega_c = 2\pi f_c = \sqrt{\frac{4}{LC}};$$

$$\text{i.e. } f_c = \frac{1}{\pi \sqrt{LC}}$$

It follows that, as drawn, the vector diagram is valid only for values of $f \leq f_c$, since $\sin \frac{\beta}{2}$ cannot be greater than 1.

f_c is called the Cut-Off Frequency.

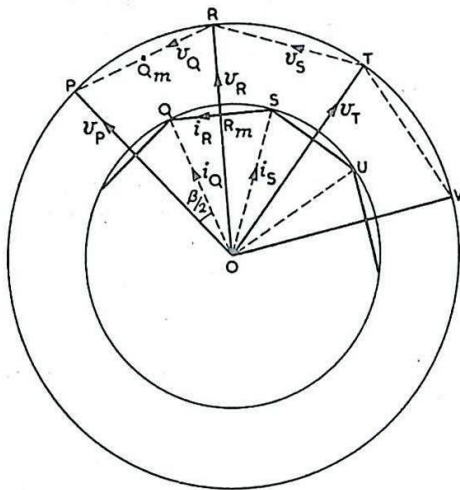


Fig. 76. - Pass-band vector diagram for low-pass filter.

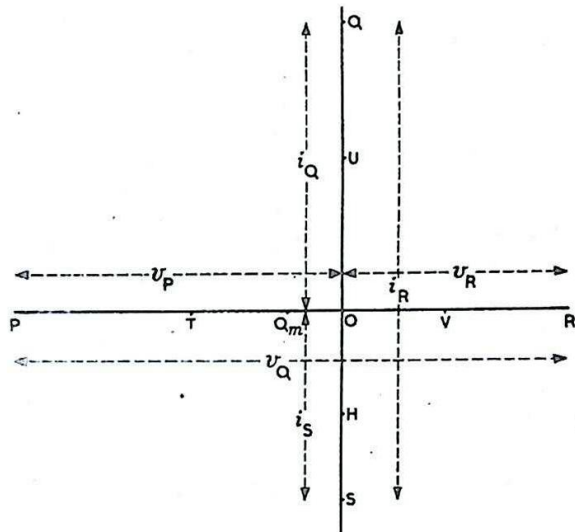


Fig. 77. - Attenuation band vector diagram for a low-pass filter.

Characteristic Impedance

The T-section characteristic impedance is the ratio of the potential difference, between Q_m and O to the current i_Q (Fig. 74(c)).

This is the ratio $\frac{Q_m O}{QO}$ (Fig. 76).

$$\begin{aligned} \text{Thus } T_{ZO} &= \frac{Q_m O}{QO} = \frac{Q_m O}{RO} \cdot \frac{RO}{QO} \\ &= \cos \frac{\beta}{2} \cdot \frac{\hat{v}_R}{\hat{i}_Q} \end{aligned}$$

But, in the triangles PRO, QSO,

$$\frac{\hat{v}_Q}{\hat{i}_R} = \frac{\hat{v}_R}{\hat{i}_Q} = \frac{\sqrt{\hat{v}_R \cdot \hat{v}_Q}}{\sqrt{\hat{i}_R \cdot \hat{i}_Q}} = \sqrt{\frac{\omega L}{\omega C}} = \sqrt{\frac{L}{C}}$$

$$\text{Hence } T_{ZO} = \sqrt{\frac{L}{C}} \cos \frac{\beta}{2}$$

Also $Q_m O$ and QO are collinear, showing that the impedance is purely resistive. (Similarly, from $R_m O$ and RO , corresponding to the π -division of Fig. 74(e), it may be shown that $\pi_{ZO} = \sqrt{\frac{L}{C}} \sec \frac{\beta}{2}$, and is purely resistive).

Attenuation Band

At frequencies higher than the cut-off frequency, the phase-shift β remains at its cut-off value, π radians per section. The vector diagrams collapse into single lines, as shown in Fig. 77, so that equation (1) in the analysis of the pass band no longer holds. All the currents are in quadrature with the corresponding voltages, and the characteristic impedance for the T-section is an inductive reactance.

Attenuation per section

Referring to Fig. 77

$$\begin{aligned}\hat{i}_R &= \hat{i}_Q + \hat{i}_S = (k+1) \hat{i}_S; \\ \hat{v}_Q &= \hat{v}_P + \hat{v}_R = (k+1) \hat{v}_R \dots\dots\dots (2)\end{aligned}$$

Hence

$$\frac{\hat{v}_Q}{(k+1) \hat{i}_S} = \frac{(k+1) \hat{v}_R}{\hat{i}_R}$$

$$\text{i.e. } \frac{k \hat{v}_S}{(k+1) \hat{i}_S} = \frac{(k+1) \hat{v}_R}{\hat{i}_R}$$

$$\text{Therefore } \frac{k \omega L}{k+1} = \frac{k+1}{\omega C}; \text{ i.e. } \frac{(k+1)^2}{k} = \omega^2 LC = \frac{4 \omega^2}{\omega_c^2}$$

$$\text{Hence } \frac{k+1}{\sqrt{k}} = \frac{2\omega}{\omega_c}$$

$$\text{or } \sqrt{k} + \frac{1}{\sqrt{k}} = \frac{2f}{f_c}$$

From this it may be deduced that the attenuation per section is given by:-

$$\sqrt{k} = \frac{f}{f_c} + \sqrt{\left(\frac{f}{f_c}\right)^2 - 1}$$

(The propagation constant, γ , is $\alpha + j\beta$, where $\beta = \pi$ and $\alpha = \log_e k$ nepers per section).

Characteristic Impedance

In the attenuation band

$$\begin{aligned}T_{X0} &= \frac{Q_{m0}}{Q_0} \\ &= \frac{\hat{v}_P - \hat{v}_R}{2 \hat{i}_Q} \\ &= \frac{\hat{v}_R (k-1)}{2 \hat{i}_Q} \\ &= \frac{\hat{v}_Q (k-1)}{(k+1) 2 \hat{i}_Q} \dots\dots\dots \text{from (2) above} \\ &= \frac{\omega L}{2} \cdot \frac{(k-1)}{(k+1)} \\ &\doteq \frac{\omega L}{2} \quad \text{for } k \text{ large;} \\ \text{i.e. for } f >> f_c.\end{aligned}$$

This gives the magnitude of the characteristic impedance.

The above results may be obtained more readily by the use of the theory of complex numbers, applied to the formulae given in Section 3.

11. Graphical Representation of Properties of a Simple Low-Pass Filter T-Section

For a network containing the ideal components of Fig. 74(a) the characteristic impedance and attenuation vary with frequency as shown in Fig. 78. These variations are obtained from the formulae derived above.

It is common practice to draw the Insertion Loss-Frequency curve, instead of the attenuation frequency curve, and the loss is plotted on an inverted axis, as shown in Fig. 78(c). The insertion loss per section is $20 \log_{10} k$ decibels. (Insertion loss usually implies voltage loss (see Sec. 5). In this case the section is properly matched so that voltage and power losses have the same value).

Effect of incorrect termination

It is not generally possible to terminate such a network by its characteristic impedance at all frequencies, because of the inconvenient form of the variations which occur in this quantity with changes of frequency. In practice, the simplest compromise available is to make the load a resistance of value $\sqrt{\frac{L}{C}}$. The filter is

then properly terminated at zero frequency, and the effect of incorrect termination is not pronounced until cut-off frequency is approached, particularly if the output impedance of the source is also a resistance of value $\sqrt{\frac{L}{C}}$. The effect is illustrated in Fig. 79, where the loss-frequency

characteristic is plotted for a single T-network terminated in $\sqrt{\frac{L}{C}}$.

The input impedance is not purely resistive except at zero frequency, and some loss (3 db. at f_c) is introduced into the pass band. If several sections are used in cascade, terminated in $\sqrt{\frac{L}{C}}$, or if a matching

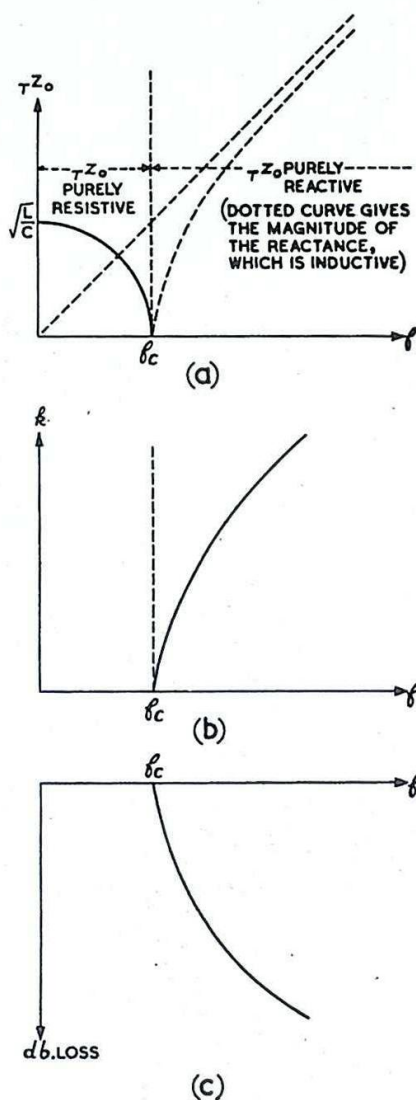


Fig. 78.- Characteristic properties of a low-pass filter (T-section).

section is inserted between the T-network and the load, the loss-frequency characteristic for each section will lie generally between the two curves of Fig. 79.

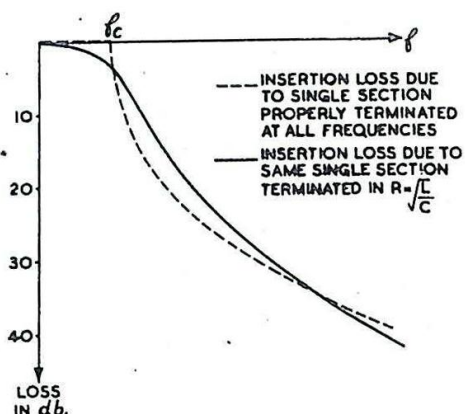


Fig. 79.- Insertion loss for single low-pass filter section terminated in $\sqrt{L/C}$ at all frequencies.

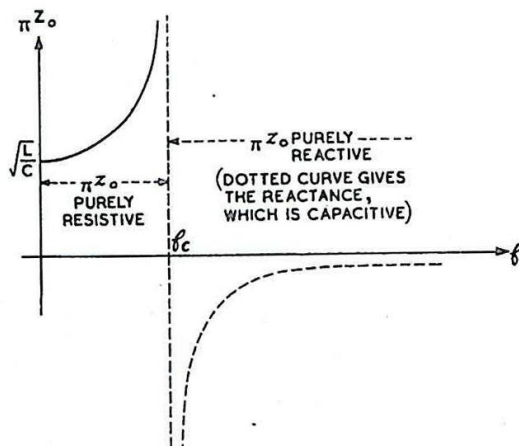


Fig. 80.- Characteristic impedance of a low-pass filter.

12. Properties of Simple Low-Pass Filter π -sections

The attenuation due to a filter formed of π -sections is the same as for the corresponding T-section network, as described in Sec. 6. The principal difference between the two types lies in their characteristic impedances, that of the π -section becoming infinite at the cut-off frequency, and becoming a capacitive reactance in the attenuation band (Fig. 80).

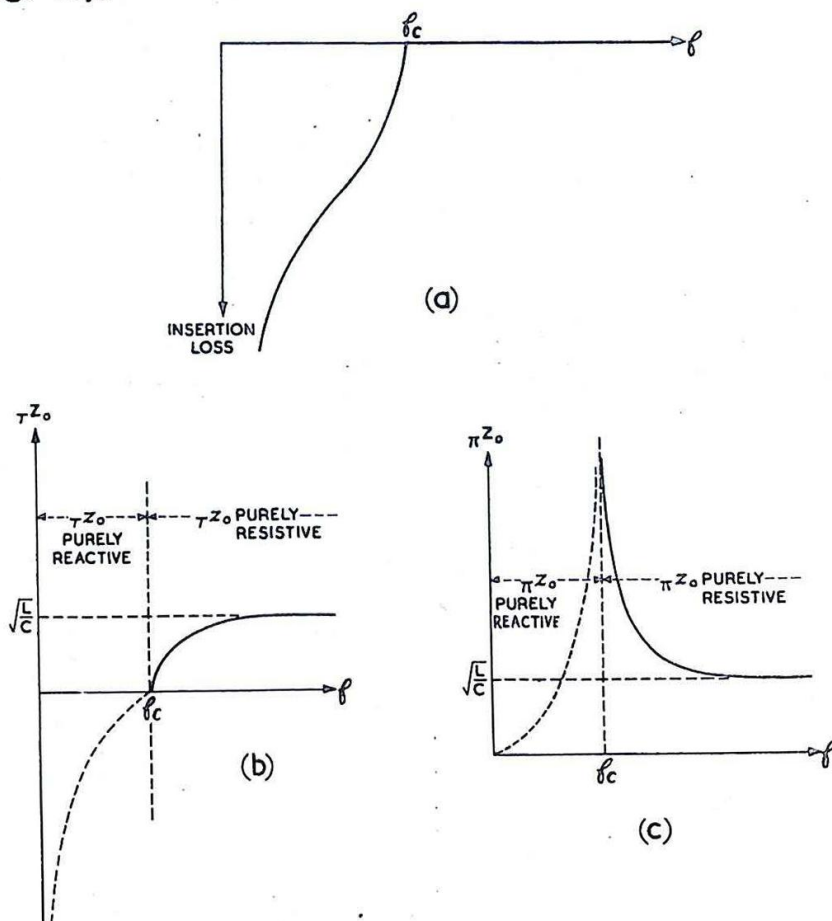


Fig. 81.- Insertion loss and characteristic impedance curves for a simple high-pass filter.

The two characteristic impedances are related by the formula

$$T Z_0 \cdot \pi Z_0 = z_1 \cdot z_2 = \frac{L}{C}$$

13. Properties of Simple High-Pass, Band-Pass and Band-stop Filters

The insertion loss curves and characteristic impedances of the simple high-pass filters of Fig. 72(c) and (d) are shown in Fig. 81.

A band-pass filter may be formed of a high-pass and a low-pass filter in cascade. Idealised insertion loss curves for such a filter are shown in Fig. 82, but these are not attainable in practice due to unavoidable resistive losses and also to the impossibility of terminating both filters correctly at all frequencies within the pass band.

Similarly a band-stop filter may be formed by bridging a high-pass and low-pass filter, as shown in Fig. 83.

The types of band-pass and band-stop filters commonly met with in radio receivers are simpler in that they involve fewer components than are required by the bridge or cascade connections just described, but the analysis and design are more complicated and the attenuation-frequency characteristics are usually much inferior to the idealised curves of Figs. 82 and 83. For RF and IF circuits of a radio receiver such filters are generally made adjustable, and the final sizes of components chosen empirically.

14. Other Types of Filters

Networks more complicated than those described above are frequently used as band-pass, band-stop and anti-carrier-wave (ACW) filters. Typical arrangements are shown in Figs. 84 to 87, with ideal attenuation characteristics. The insertion of a parallel anti-resonant circuit in the series arm, or a series resonant circuit in the shunt arm introduces infinite attenuation at the resonant frequency of the tuned circuit.

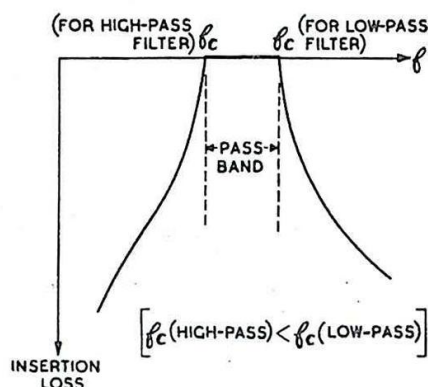


Fig.82.- Insertion loss for band-pass filter formed by high-pass and low-pass filters in cascade.

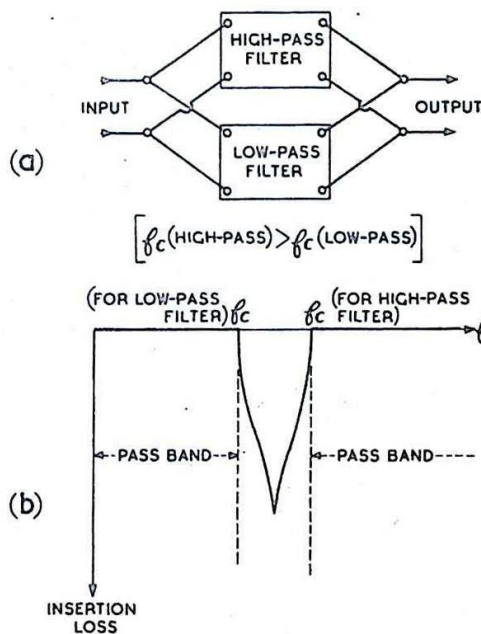


Fig.83.- Band-stop filter formed by bridging high-pass and low-pass filter.

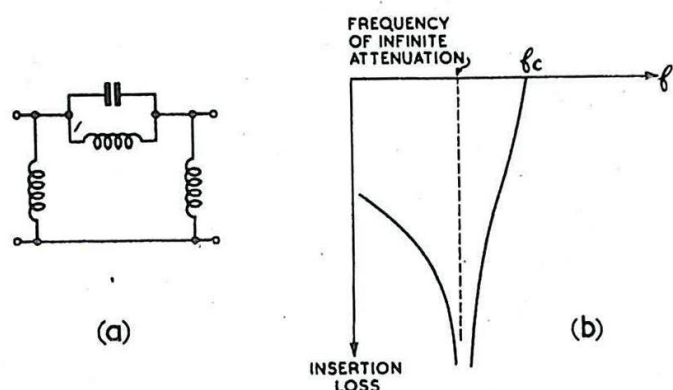


Fig.84.- High-pass m-derived π -section filter.

In practice, resistive losses modify the idealised curves, introducing a certain amount of attenuation in the pass-band, and generally decreasing the sharpness of the distinction between pass and attenuation bands. High-Q circuits and components are required if the ideal curves are to be approached closely; for this reason piezo-electric crystals are frequently used in place of tuned circuits in band-stop and band-pass filters, but their use is limited by their very narrow tuning range.

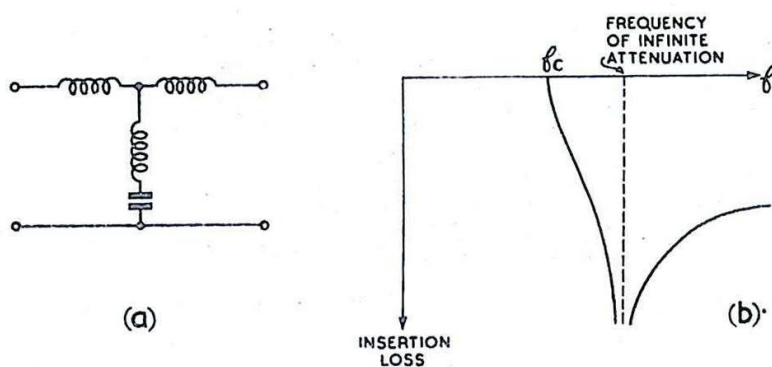


Fig.85.- Low-pass m-derived T-section filter.

15. Delay Networks

It is shown in Chap. 4. Sec. 6 how the change of phase produced in a transmission line depends on the time taken for an electromagnetic wave to pass along the line. The same is true for an artificial line or Delay Network. If this transit time is T , then the phase angle ϕ , by which the output lags the input, is connected with T and f , the frequency, by the relation

$$\phi = 2\pi fT$$

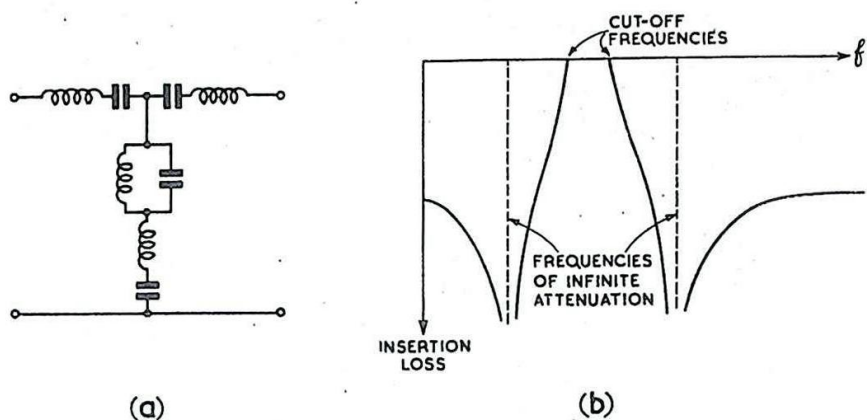


Fig.86.- Band-pass T-section filter.

For a simple low-pass network of n sections properly terminated, the time of transit for a wave of frequency $f \ll f_c$ may be taken as

$$T = \frac{\theta}{2\pi f} = \frac{n\beta}{2\pi f}, \quad \text{where } \sin \frac{\beta}{2} = \frac{f}{f_c} = \pi f \sqrt{LC}.$$

Hence, for $f \ll f_c$, when $\sin \frac{\beta}{2} \doteq \frac{\beta}{2}$,

$$T \doteq \frac{n}{2\pi f} \cdot 2\pi f \sqrt{LC}$$

$$\doteq n \sqrt{LC}. \quad (\text{This result is independent of } f, \text{ provided}$$

$f \ll f_c$.)

For higher values of f (but still within the pass-band), the value of T increases, until at $f = f_c$ the transit time is $T = \frac{n\pi}{2\pi f_c} = \frac{n}{2f_c} = \frac{n\pi}{2} \cdot \sqrt{LC}$. The variation of β and T with frequency

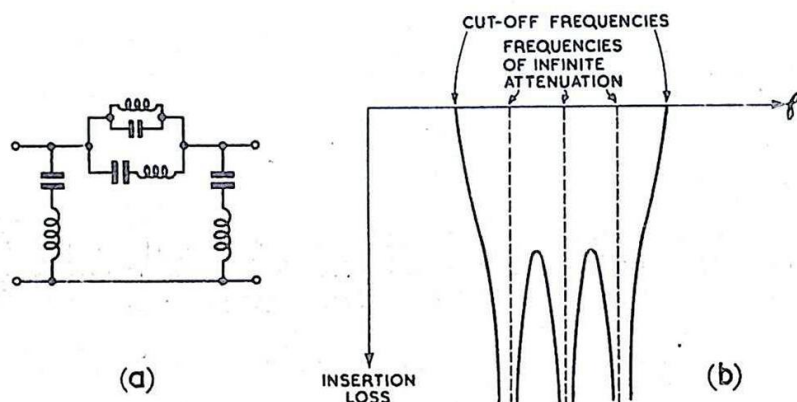


Fig.87.- Band-stop π -section filter.

is illustrated in Fig. 88.

This increase in time-delay with rising frequency causes distortion of non-sinusoidal waves, this distortion being superimposed on that due to the attenuation of components whose frequency is higher than the cut-off frequency. Where such a network is used to delay a rectangular pulse the cut-off frequency should be high compared with $\frac{1}{T_p}$, where T_p is the pulse width. This determines the maximum permissible value of \sqrt{LC} , so that the number of sections necessary may be determined from the formula:-

$$n = \frac{T}{\sqrt{LC}}, \quad T \text{ being the total delay required.}$$

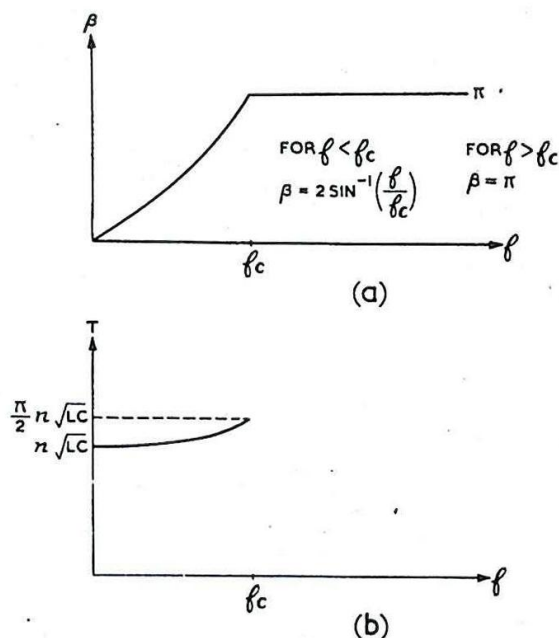


Fig. 88.- Variation of phase shift and delay time with frequency.

The actual values of L and C needed are determined from the chosen cut-off frequency and the value R_ℓ of the resistive termination of the network. Thus :-

$$f_c = \frac{1}{\pi \sqrt{LC}} \quad \text{and} \quad R_\ell = \sqrt{\frac{L}{C}},$$

$$\text{whence } C = \frac{1}{\pi f_c R_\ell} \text{ farads, and } L = \frac{R_\ell}{\pi f_c} \text{ henries,}$$

R_ℓ being measured in ohms and f_c in c/s.

16. Use of Delay Line for Producing Rectangular Pulses

It is shown in Chap. 4, Sec. 12 that a length of transmission line may be used to produce narrow, rectangular pulses. The advantage of using a uniform cable for such purposes is that it passes all waveforms without distortion. As described in Chap. 4, the pulse is produced by a wave travelling to the end of the cable and back again, so that the cable transit time must be half the desired pulse width.

An artificial line formed of from two to eight low-pass filter sections may be used in place of the cable, with considerable saving of both space and weight.

Since $T = \frac{1}{2} T_p$, the design requirements for the network given in Sec. 15 include the relations

$$\frac{1}{\pi \sqrt{LC}} \gg \frac{1}{T_p} \quad \text{and} \quad \frac{1}{2} T_p = n \cdot \sqrt{LC}$$

$$\text{so that} \quad \frac{1}{\pi \sqrt{LC}} \gg \frac{1}{2 n \sqrt{LC}}$$

$$\text{or } n \gg \frac{\pi}{2}.$$

For values of n greater than 6, or thereabouts, the pulse can be made substantially rectangular.

Examples

- (i) To construct a low-pass filter T-section whose characteristic impedance at zero frequency is 600 ohms and whose cut-off frequency is 4 Mc/s.

$$\text{We have } \sqrt{\frac{L}{C}} = 600 \quad \text{and} \quad \frac{1}{\pi\sqrt{LC}} = 4 \cdot 10^6$$

$$\text{Hence } \sqrt{LC} = \frac{1}{4\pi \cdot 10^6}$$

$$\therefore L = \frac{600}{4\pi \cdot 10^6}$$

$$= \frac{150}{\pi} \text{ microhenries}$$

$$\doteq 47.7 \text{ microhenries.}$$

$$\text{Similarly } C = \frac{1}{4\pi \cdot 10^6} \cdot \frac{1}{600}$$

$$= \frac{1}{24\pi \cdot 10^8}$$

$$\doteq 133 \text{ pF.}$$

Hence the T-section has two series inductances each of 24 microhenries and a shunt capacitance of 133 pF.

- (ii) To construct a delay line to pass signals of all frequencies up to 3 Mc/s and to provide a total delay of 6 μ s. It must match to a 2000 ohm line.

$$\text{We make } f_c = \frac{1}{\pi\sqrt{LC}} = 3 \cdot 10^6 \text{ and } \sqrt{\frac{L}{C}} = 2000.$$

As for the previous example we obtain

$$\sqrt{LC} = \frac{1}{3\pi \cdot 10^6}$$

$$\text{Whence we obtain } L = 212 \mu\text{H.}$$

$$C = 53 \text{ pF.}$$

The total delay is 6 μ s, whilst the delay per section is

$$\sqrt{LC} = \frac{1}{3\pi} \mu\text{s.}$$

Hence the number of sections required is

$$6 \div \frac{1}{3\pi} = 18\pi \doteq 57$$

Hence 57 sections are required, each with total series inductance $212 \mu\text{H}$ and total shunt capacitance 53 pF .

- (iii) To construct a pulse-forming network of 6 sections to match to a 800 ohm load and to have a double transit-time of $1.0 \mu\text{s}$.

$$\text{We have } \sqrt{\frac{L}{C}} = 800.$$

$$6 \sqrt{LC} = 0.5 \cdot 10^{-6}$$

$$\text{Hence } 6L = 800 \cdot 0.5 \cdot 10^{-6}$$

$$\therefore L = \frac{400}{6} \text{ microhenries}$$

$$\approx 67 \text{ microhenries.}$$

$$\text{Similarly } 6C = \frac{0.5 \cdot 10^{-6}}{800}$$

$$\text{Hence } C = \frac{0.5 \cdot 10^6}{6 \cdot 800} \text{ pF}$$

$$\approx 104 \text{ pF.}$$

Hence each section has total series inductance $67 \mu\text{H}$ and shunt capacitance 104 pF .

RESOLVING NETWORKS

17. General

It is often required in radar control or generating circuits to produce a voltage proportional to the sine or cosine of some variable angle. This angle may be an angle of azimuth or elevation, or it may be a phase angle corresponding to a time-delay, such as is involved in the measurement of range. In any case there are three principal types of resolver that are used in practice; these are :-

- (i) the Resistive Resolver (Sine Potentiometer),
- (ii) the Inductive Resolver (Magalip Resolver),
- (iii) The Capacitive Resolver.

Type (i) may be used with DC supplies; types (ii) and (iii) can be used only with AC supplies. Type (iii) is not encountered as a resolver alone, being employed only in the capacity phase-shifting network described in Sec. 22. It will not be described separately.

18. Resistive Resolver

The basis of this device is illustrated in Fig. 89. A wire-wound resistor OP is made into a non-linear potentiometer so that for movement θ of the slider the resistance tapped between O and Q is proportional to $\sin \theta$.

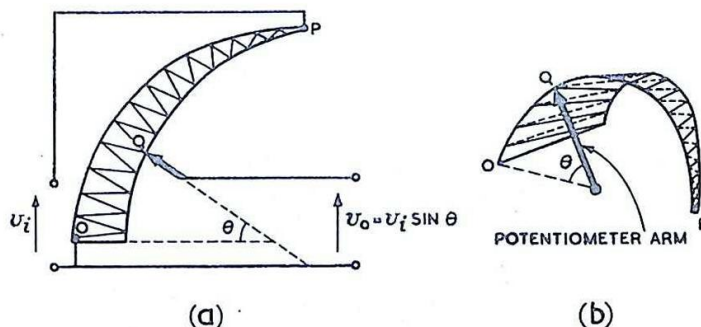


Fig. 89. - Resistive resolver (sine potentiometer).

Since $R \propto \sin \theta$

$$\delta R \propto \cos \theta \delta \theta,$$

so that if δR is the incremental resistance for each turn of the wire on the former, corresponding to a length δl , and provided the resistance per unit length of the wire is uniform,

$$\delta l \propto \cos \theta.$$

This means that the wire may be wound on a former having the shape of a cosine function as shown (Fig. 89(b)).

A group of four such cosine-law resistors may be arranged as shown in Fig. 90(a) to give continuous resolution. As the potentiometer arm turns through an angle θ the direct voltage tapped from the resolver varies as shown at (b). If an alternating supply is used, as shown in

Fig. 91 the input voltage being v_i , the output voltage is $v_o = v_i \sin \theta$, as for the DC circuit.

Such a resolver, although accurate for direct or low-frequency alternating currents, is not often used for frequencies of several kilocycles or more because of the effect of stray capacitance, and because inductive resolvers are more suitable at such frequencies. Instead of the accurately wound sinusoidal potentiometers, linear resistances may be used to give a rough type of resolution as shown in Fig. 92. The eight feed points are taken to the supply voltages as shown at (a) and as θ varies the voltage tapped from the resolver follows the graph shown at (b). Although such a device is inaccurate as a straightforward resolver, it finds a useful application in the resistive type of phase-shift network described in Sec. 22, where the inaccuracies of the method do not matter greatly because of the special requirements of the system.

19. Inductive Resolver (Fig. 93)

The Inductive, or Magalip, type of resolver consists fundamentally of two coils, a stator and a rotor, the alternating supply normally being fed to the former and the output being taken from the latter via sliprings. Alternatively the roles of the two coils may be reversed. The coils are mounted on a common diameter and wound in such a manner that the mutual inductance between them is proportional to the cosine of the angle θ between the coil axes; (See Chap. 18 Sec. 9) Hence the voltage induced in the output coil is proportional to the product of the input voltage and $\cos \theta$; Fig. 93(b).

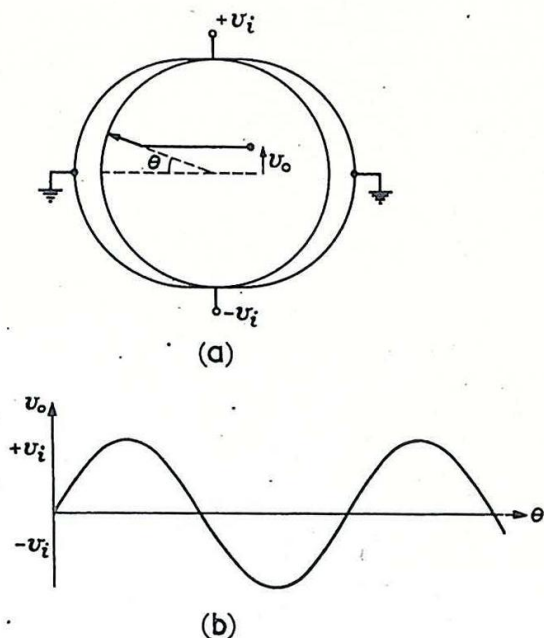


Fig. 90.- Sine potentiometer arranged to give continuous resolution.

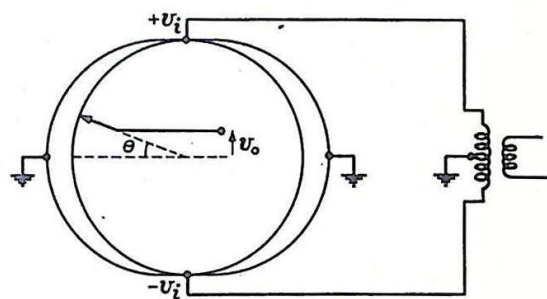


Fig. 91.- Sine potentiometers used with AC supply.

There is a large leakage inductance, so that there is very little effect of the output circuit on the input, the reflected impedance being small compared with the impedance of the primary circuit. This is important, since otherwise the variation of the reflected impedance with the position of the rotor would appreciably affect the primary current and introduce errors whose magnitude would vary with the secondary load.

If the supply is fed to the rotor, and two identical stators are used, perpendicular to each other, voltages proportional to $\sin \theta$ and $\cos \theta$ may be obtained simultaneously from the two stators, as indicated at (c).

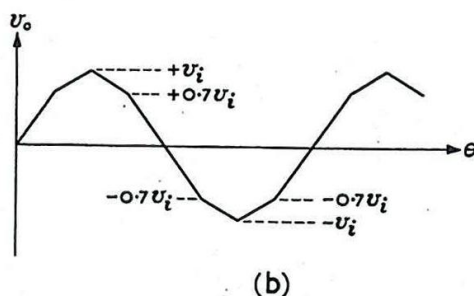
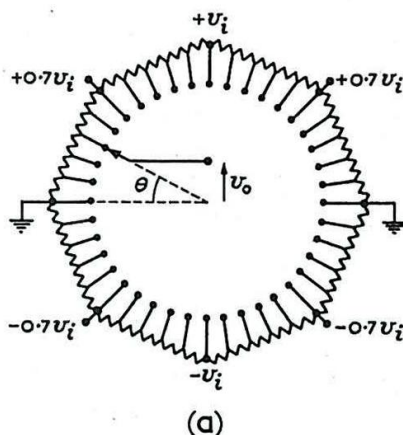


Fig.92.- Approximate resolution using linear potentiometers.

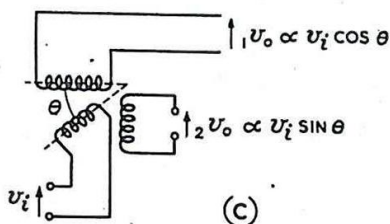
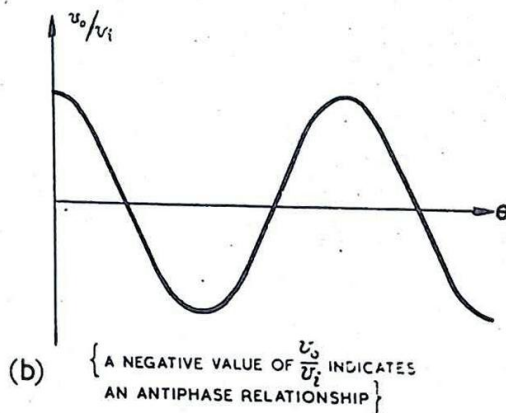
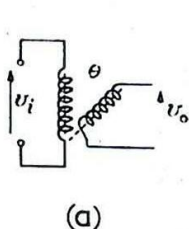


Fig.93.- Inductive resolver.

CONTINUOUSLY VARIABLE PHASE-SHIFTING
DEVICES

20. General

In sections 8 - 11 it was shown how, in an iterative filter network, there is a progressively increasing phase delay between output and input terminals as the number of sections is increased. If the number of sections is varied it is possible to choose a voltage with any desired phase relationship to the input, within the limits of the phase-step from one section to the next. In radar it is often necessary to be able to perform this process of phase variation throughout the range $0 - 360^\circ$ continuously, for the purpose of obtaining a sinusoidal voltage variation in the required phase relation with a fixed-frequency input voltage. This process is not concerned with time-delays and should not be confused with the more general problem of the use of delay lines, usually employed to introduce time-delays into circuits handling signals of pulse form; although the latter method is applicable to the production of phase delays in sinusoidal oscillations, the circuits are usually more clumsy and of wider application than is necessary. For example, where this method is employed, for the sake of accuracy the phase-change per section must be kept small, so that a large number of sections must be used.

The more common method of providing $0 - 360^\circ$ continuously variable phase-shift is to start with two sinusoidal voltages of the required frequency, in quadrature. If these are added or subtracted in the correct proportions an output voltage may be obtained with any desired phase relation to the input.

Thus, suppose the input voltages are given by

$$v_1 = A \sin \omega t \text{ and } v_2 = A \cos \omega t$$

respectively. If a fraction $\sin \beta$ of the second is subtracted from a fraction $\cos \beta$ of the first, an output voltage is provided given by

$$\begin{aligned} v_o &= A \sin \omega t \cdot \cos \beta - A \cos \omega t \cdot \sin \beta \\ &= A \sin (\omega t - \beta) \end{aligned}$$

i.e., a signal delayed by a phase angle β with respect to the input voltage v_1 .

A description is given below of three types of network used to provide continuously variable phase-shift by this method using inductive, capacitive and resistive resolvers respectively.

The Goniometer (Inductive Phase-Shifting Network) and the Capacitive Phase-Shifting Network both provide smoothly variable outputs, rely for their accuracy on manufacturing tolerances, and are not normally adjustable (save, perhaps, for an overall adjustment equivalent to ensuring that the input voltages v_1 and v_2 are of equal amplitude and truly in quadrature). The Potentiometer Phase-Shifting Network is a step-by-step device relying for its accuracy on the tolerances of its component resistors; any of these may be changed without upsetting the mechanical construction of the device. By the use of a sufficiently large number of resistive steps the errors due to the device not being smoothly variable may be made negligible.

The potentiometer circuit is substantially independent of frequency. This is not usually true of the two other circuits.

21. The Goniometer

This device uses inductive resolvers. The two quadrature input voltages v_1 and v_2 are applied to two stator coils whose axes intersect symmetrically at right angles. A rotor coil is assembled so that its axis is in the plane of the axes of the stators and may be rotated into coincidence with either. This is illustrated schematically in Fig. 94. Normally the stator coils are split into pairs in order to generate a symmetrical stator flux, and opposite coils are connected in series.

A method of correctly feeding such a four-stator goniometer from a single-phase supply is shown in Fig. 95. The simplified circuit is given at (b). In this circuit L , the inductance of a pair of coils in series, is the same for each stator group. If R_1 , R_2 and C are correctly adjusted the currents in the stators are equal in magnitude and are in quadrature. Such an arrangement is highly sensitive to changes of frequency. For a single frequency the goniometer can be made accurate to a fraction of a degree.

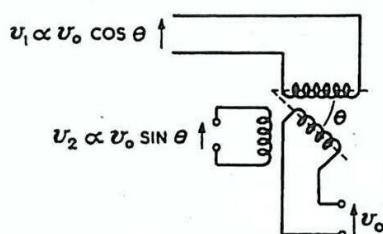
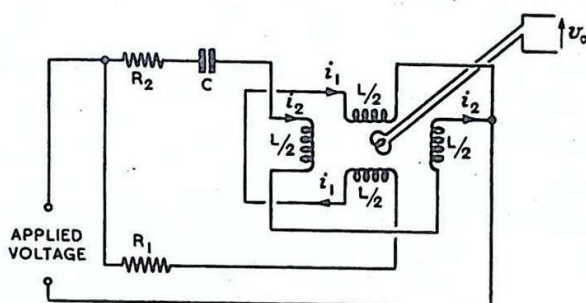


Fig. 94. - Goniometer.

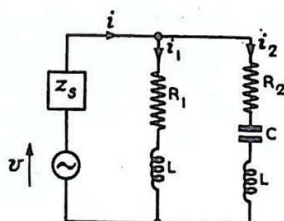
22. The Capacity Phase-Shifter (Fig. 96)

This device is closely analogous to the goniometer, relying for its operation on induced electric, instead of induced magnetic fields. It takes the form of a split-stator condenser, the lower stator plate being earthed and the upper stator plate being divided into four equal sectors. Opposite pairs of plates are fed with the balanced quadrature input voltages v_1 and v_2 , and an unbalanced output is taken between the rotor and the earthed lower plate. If the stator and rotor plates are made to the correct shapes reasonably accurate resolution is obtained by this method, errors being normally of the order of one degree.

The method of feeding the input voltages in the correct phase from a single-phase supply is the same as for the potentiometer arrangement given below.



(a) SCHEMATIC DIAGRAM OF A GONIOMETER (4 STATORS)



(b) SIMPLIFIED ELECTRICAL CIRCUIT OF (a)

Fig. 95. - Feeding a four-stator goniometer.

23. The Potentiometer Phase-Shifting Network

This device incorporates the approximate type of resolver described in Sec. 18.

In theory, accurate resolution could be obtained using tapered resistors as described in Sec. 17, and arranged as indicated in Fig. 97. Balanced sinusoidal input voltages v_1 , v_2 are applied in quadrature as shown, and the balanced output voltage is obtained from the two tapping points. In practice the necessity for accurate winding of the sinusoidal potentiometers is a drawback and linear potentiometer arrangements are used. In these a slight variation in output amplitude with phase is inevitable, but this is not usually important.

Fig. 98 gives the schematic layout of a 4-feed-point phase-shifting network, whilst the vector diagram of Fig. 99 illustrates the action. As P, Q move from A to B and from C to D respectively, the output voltage vector $P'Q'$ changes from $A'C'$ to $B'D'$, equal resistance changes corresponding to equal movements along the sides $A'B'$, $C'D'$ of the square. It follows that if the resistances between adjacent studs are all equal, equal movements of the output arm PQ do not correspond to equal increments of phase β . The graph showing errors due to using equal resistances in this four-feed-point network is shown in Fig. 100. To avoid the use of resistors of different values

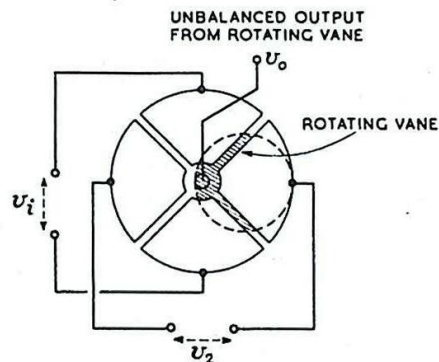


Fig. 96. - Capacitive phase-shifting network.

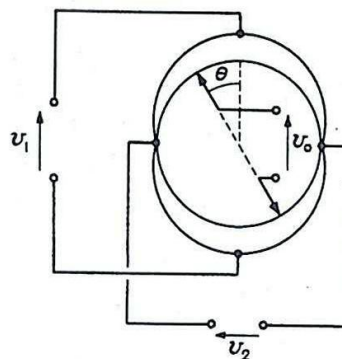


Fig. 97. - Resistive phase-shifting network, using sine-potentiometers.

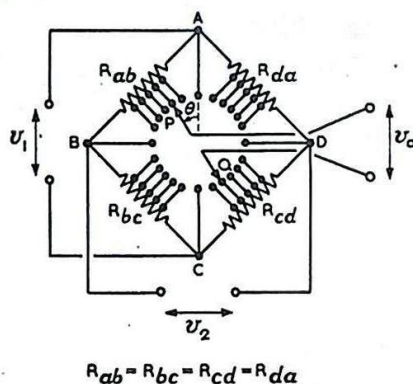


Fig. 98. - 4-feed-point potentiometer phase-shifting circuit.

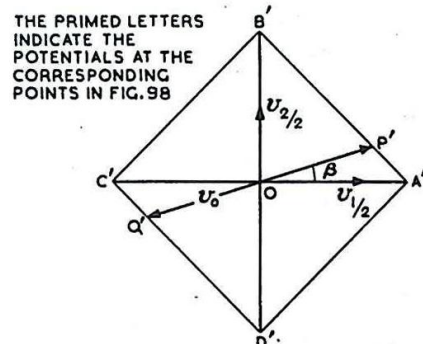


Fig. 99. - Variation of output voltage with movement of P.

the eight-feed-point arrangement shown in Fig. 101 may be employed. If the feed resistors between the input terminals and the circular resistor chain are suitably chosen the vector diagram takes the form given in Fig. 102. The output phase β is then the same as the angular rotation θ at 16 points instead of 8, as in the four-feed-point arrangement, and the intermediate errors are correspondingly reduced (Fig. 103).

One method of feeding the four points (ABCD) of Fig. 98 is shown in Fig. 104. A circuit similar to that of Fig. 104 may also be used to supply the four primary feed-points (ABCD) of Fig. 101. If the input resistance of the potentiometer between A and B or any corresponding pair

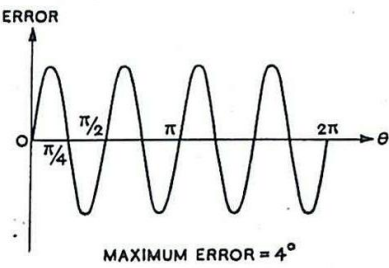


Fig. 100. - Error curve for 4-feed-point potentiometer.

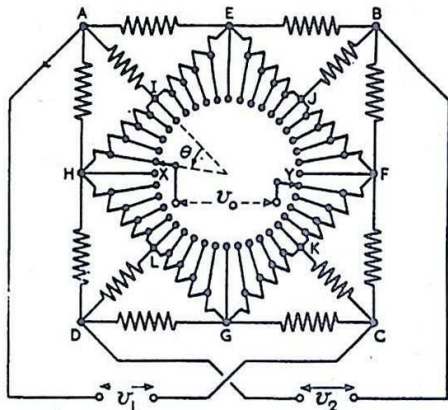
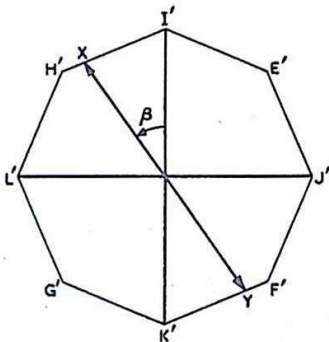


Fig. 101. - 8-feed-point potentiometer phase-shifting circuit.



THE PRIMED LETTERS INDICATE THE POTENTIALS AT THE CORRESPONDING POINTS IN FIG. 101

Fig. 102. - 8-feed-point potentiometer circuit; variation of output phase and amplitude with slider movement.

of points is R_p then unless

$$R_p \gg R \dots\dots\dots (1)$$

the phase-shift circuit will be appreciably affected.

If this inequality (1) is satisfied, then the condition for voltages v_1 and v_2 to be in quadrature is

$$\omega CR = 1,$$

since then the voltage at B lags that at A, whilst the voltage at D leads that at A, both by 45° , as indicated in the vector diagram of Fig. 104(c). The voltages v_1 (between A and C) and v_2 (between B and D) are then equal and in quadrature.

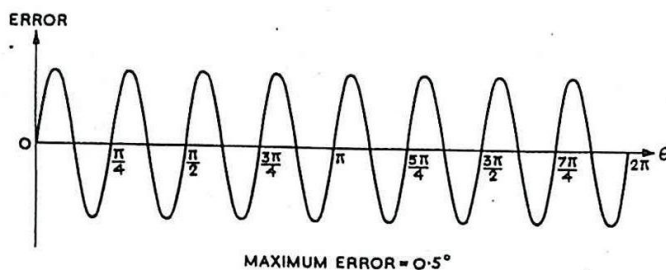


Fig. 103.- Error curve for 8-feed-point potentiometer circuit.

If the input resistance R_p between each pair of the feedpoints of the potentiometer is not negligible, the error thereby introduced can be minimised by inserting a small coil L , shown dotted in Fig. 104(a), in series with each resistor R . It can be shown from the geometry of the vector diagram (d) or by any other method of analysis that if the relations

$$L = \frac{R}{\omega^2 C (R + R_p)}$$

and

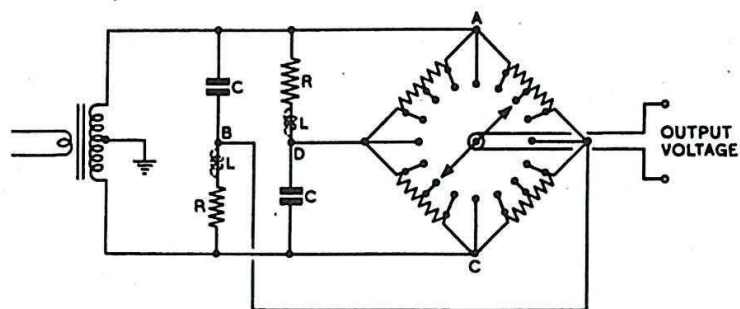
$$\frac{\omega C R R_p}{R + R_p} = 1$$

are satisfied, the figure formed by the points $A'B'C'$ and D' of Fig. 104(d) is very nearly a perfect square, so that v_1 and v_2 are again equal and in quadrature.

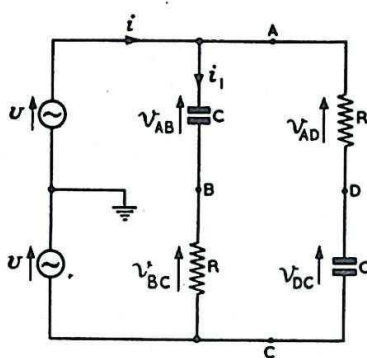
Alternatively, the input resistance of the potentiometer circuit can be made ineffective by feeding the voltages via cathode follower circuits, which owing to their large input impedance do not appreciably load the phase-shifting C-R network.

24. 0 - 180° Phase Shifting Network

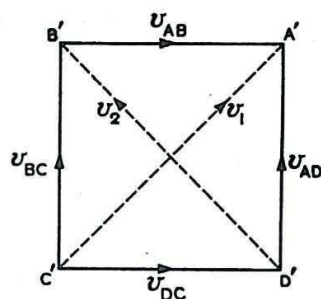
The method described in Sec. 23 of feeding the four points from a single-phase supply may readily be adapted to provide variable phase-shift from 0 to 180° only. The circuit arrangement is indicated in Fig. 105(a), and the vector diagram (b) shows the action of the network. Provided the output impedance of the transformer is small compared with the input impedance of the network, the voltage developed between X and Y is constant, and is represented by $X'Y'$ in (b). $X'Z'$ and $Y'Z'$ represent the quadrature voltages developed across C and R respectively.



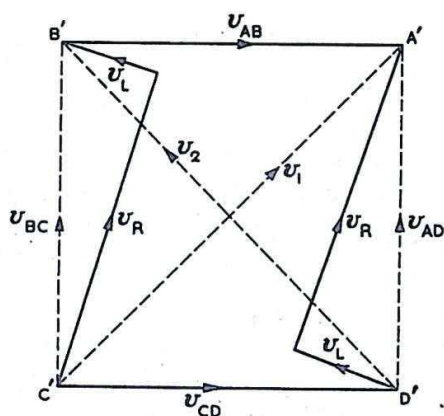
(a) SCHEMATIC DIAGRAM OF QUADRATURE FEED NETWORK FOR POTENTIOMETER PHASE SHIFTING CIRCUIT



(b) SIMPLIFIED ELECTRICAL CIRCUIT OF FIG (a)



(c) VECTOR DIAGRAM



(d) EFFECT OF INTRODUCING SERIES INDUCTANCE L TO COMPENSATE FOR RESISTANCE R_p SHUNTING C

Fig. 104. - Quadrature feed network for potentiometer phase-shifting circuit.

The vector $O'Z'$ represents the output voltage in magnitude and phase. As the value of R is varied the point Z' moves on a semi-circle on $X'Y'$ as diameter with centre O' , since angle $X'Z'Y'$ is always a right-angle. The output voltage, represented by $O'Z'$, therefore remains constant in magnitude, as the value of R (or C) is varied, and its phase relative to that of the voltage across the transformer secondary winding can be changed through about 140° with practical values of R and C .

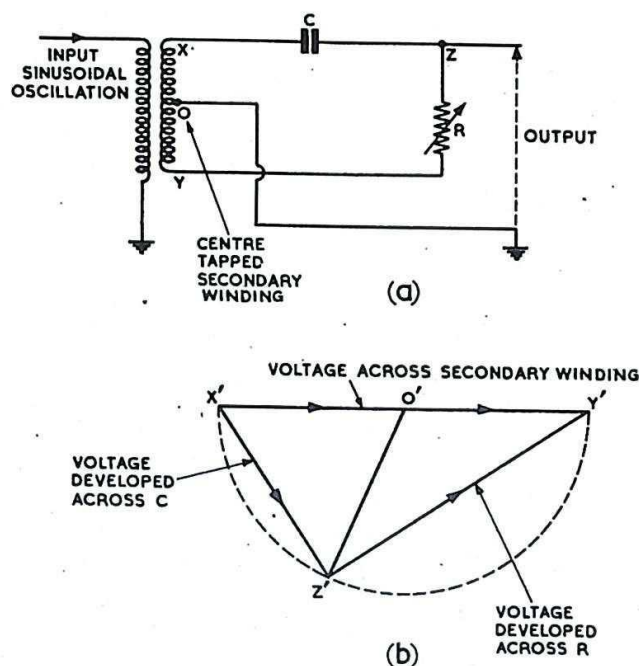


Fig. 105. - 180° phase-shifting network.

CHAPTER 1

TRANSMISSION LINES

INTRODUCTION

1. Circuits With Distributed Constants

The theory of circuits with "lumped constants", based on the laws of Kirchhoff, Ohm and Faraday, is really the theory of idealised circuits having no "size", i.e., in which all significant dimensions are small compared with the appropriate wavelength. Although the components in the material arrangements have size and shape, with electrical and magnetic properties which involve the space around them, simple circuit theory reduces these complex properties to those of ideal resistances, inductances and capacitances, in circuits where there is no time delay between the application of an EMF at one point and the related voltage produced at another. A simple example is the tendency, in circuit theory, to consider only the components, or lumped constants, and to ignore the connections or "short-circuits" which form, in low-frequency circuits, the greater portion of the path-length of a complete network. The interactions between the currents in these connections, and the time taken for a wave to travel along them, are of major importance at high frequencies.

The theory of circuits having "distributed constants" makes an attempt to widen the scope of the method of lumped constants, with considerable success. However, particularly at very high frequencies, there are many problems to which it fails to provide a satisfactory answer, and it is often necessary to resort to the fundamental physics of electromagnetism.

In this consideration of transmission lines, the transition from the theory of lumped constants to that of distributed constants will be preceded by a discussion of electromagnetic wave propagation formulated mainly in terms of the current in, and potential differences between, the conductors which form the lines. The more fundamental considerations of the behaviour of the electric and magnetic fields associated with such currents and voltages will be dealt with in Chap. 5.

2. Transmission Lines

In general terms, a Transmission Line is a device for guiding the flow of electromagnetic energy from one place, where it is available, to another, where it is to be utilised. The word Wave-guide has precisely the same significance, though it is usual to restrict the application of this term to systems embodying a single conductor (or dielectric rod, in some cases) and to use the term Transmission Line for systems using two conductors (or more).

A uniform transmission line is one in which the electrical properties of the line (per unit length) do not vary throughout its length; i.e. it usually has a constant cross-section. Non-uniform, or tapered, transmission lines are occasionally used, but their properties are not covered by this chapter, which is concerned solely with the behaviour of the uniform transmission line with two conductors. These conductors may be separate (parallel or twisted pair) or concentric (coaxial line). The separation between the conductors must be small compared with the wavelength. (see Secs. 7 and 46 and Chap. 5, Sec. 12)

Typical transmission lines used in radar are shown in Fig. 106. The electrical properties of the lines depend not only upon the size, shape and disposition of the conductors, but also upon the dielectric which maintains that disposition, and the extent to which the arrangement is constant, or varies with movement or change in atmospheric conditions. Where a uniform solid dielectric is used, with the wires embedded, as shown in Fig. 106(a), it is possible to maintain reasonably constant electrical properties by the exclusion of atmospheric variations and by the constant spacing which can thus be maintained between the conductors. The disadvantage of this type of cable is loss of energy in the imperfect dielectric. The other extreme is the open-wire feeder construction of Fig. 106(b). Here dielectric losses are low except at the spacers, which are apt to cause serious trouble if they get very wet.

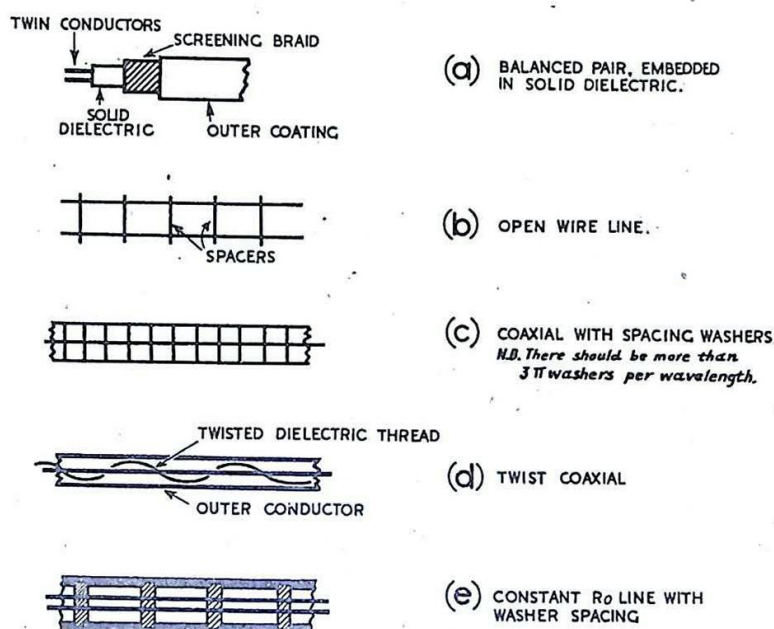


Fig. 106 - Typical transmission lines used in radar.

Their presence inevitably introduces non-uniformity, which makes the behaviour of the line more difficult to predict and thus less easy to control. This problem of suspension also applies to coaxial cables with mainly air dielectric where the inner must be maintained in a position fixed relative to the outer. Various arrangements are illustrated in Figs. 106(c), (d), (e), showing how attempts are made to satisfy both requirements, uniformity and low losses. Particular arrangements which avoid the use of dielectric supports are described in Sec. 29. Losses due to induction and radiation, which may make open-wire lines unuseable as transmission systems at VHF, is dealt with in Sec. 46.

Under certain circumstances, waveguides are to be preferred to twin-conductor systems, as discussed in Chap. 5, Sec. 5.

Although the primary function of a transmission line is to transmit energy, other applications arise, particularly with short lengths of line; e.g., reactive circuit elements, resonant circuits, and impedance-transforming devices. In these special applications, as with all reactors, it is the storage, rather than the transmission

of energy, which is more important.

TRAVELLING WAVES ON UNIFORM LOSS-FREE TRANSMISSION LINES

3. Application of Circuit Laws to Transmission Lines

For simplicity it will be assumed that in the ideal transmission lines about to be discussed there are no energy losses. Such conditions approximate to those encountered in the practical cables shown in Fig. 106, or in parallel lines in the form of broad metal ribbons, pictured in Fig. 107. Since the distributions of charge and electric and magnetic fields in the latter case are more simply represented, the diagrams which follow will apply to this arrangement.

The evolution of a coaxial line from the ribbons is suggested in Fig. 108, the ribbons being bent to form the inner and outer of a coaxial pair; the electric and magnetic fields associated with the current in the conductors are then confined to the space between them, as they would be, substantially, between the parallel plates in the original arrangement.

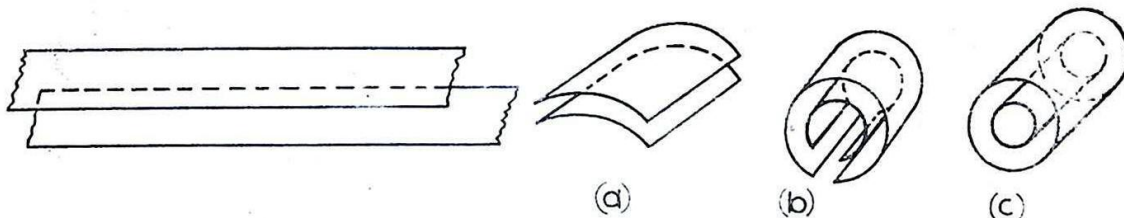


Fig. 107 - Parallel lines in form of broad metal ribbons.

Fig. 108 - Bending parallel plate lines to form coaxial pair.

It can be shown that electromagnetic waves may be propagated without distortion along such a transmission line with a velocity u where

$$u = \frac{1}{\sqrt{L_\ell C_\ell}}, \quad L_\ell \text{ and } C_\ell \text{ being the}$$

inductance and capacitance per unit length of the line.

Moreover, at any portion of the wave, a change in potential difference between the conductors is always related to a change in the wave current flowing in the conductors by the relation

$$\frac{dv}{di} = \sqrt{\frac{L_\ell}{C_\ell}}.$$

A method of deriving these fundamental results is given below. A simpler approach is given in Section 11.

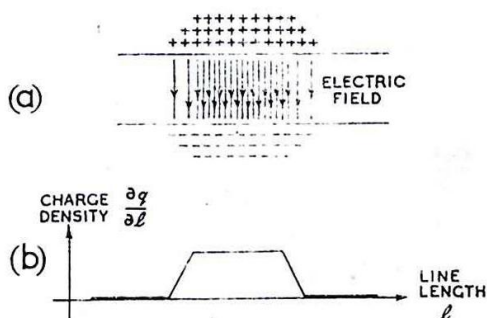


Fig. 109 - Instantaneous charge distribution on a section of the line of fig. 107.

Derivation of Formulae

Suppose there is at any instant a charge distribution on the line as illustrated in Fig. 109. This corresponds to the potential difference between the conductors shown in Fig. 110 and enlarged in Fig. 111, since $V \propto Q$. Considering the small section of the line of width δl in Fig. 112(a), it is seen that there is a resultant potential difference in the circuit equal to δv (see Fig. 112(b)). Since there are no other sources of EMF present, this must be due to the EMF induced by the changing current according to Faraday's law,

* $\delta v = -\delta L \frac{\partial i}{\partial t}$, where δL is the inductance of the small circuit.

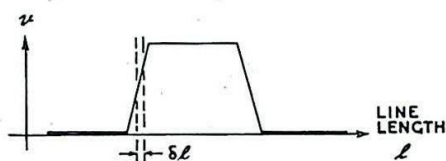


Fig. 110 - Instantaneous potential distribution corresponding to fig. 109.

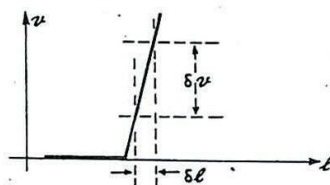


Fig. 111 - Enlarged diagram showing potential distribution of section δl .

If L_l is the inductance per unit length of the transmission line,

$$\delta L = L_l \delta l$$

so that

$$\delta v = -L_l \frac{\partial i}{\partial t} \delta l$$

i.e.

$$* \quad \frac{\partial v}{\partial l} = -L_l \frac{\partial i}{\partial t} \dots\dots\dots(1).$$

There is thus present at the sloping portion of the 'pulse' a current changing with time, so that the pulse cannot be stationary. If δi is the difference in current between the two ends of the section, (see Fig. 112(c)),

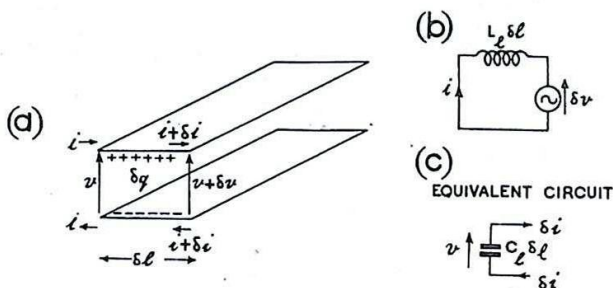


Fig. 112 - Enlarged diagram of section of line of length δl , with currents and voltages.

* $\frac{\partial}{\partial l}$ indicates a change with length l for t constant.
 $\frac{\partial}{\partial t}$ indicates a change with time t for l constant.

$$\delta i = - \frac{\partial}{\partial t} (\delta q)$$

$$= - \frac{\partial}{\partial t} (C_l \delta \ell v)$$

$$= -C_l \frac{\partial v}{\partial t} \delta \ell.$$

(If the current increases the charge left on the line decreases; hence the - sign)

Hence $\frac{\partial i}{\partial \ell} = -C_l \frac{\partial v}{\partial t} \dots\dots\dots(2).$

We shall now show that relations (1) and (2) are consistent with the pulse travelling without changing its shape, either to the right or to the left, with velocity $\frac{d\ell}{dt} = u = \frac{1}{\sqrt{L_l C_l}}$, and such

that $\frac{dv}{di} = \sqrt{\frac{L_l}{C_l}}$

From (1) we have:- $\frac{\partial v}{\partial \ell} = -L_l \frac{\partial i}{\partial t}$
 $= -L_l \left[-\frac{\partial i}{\partial \ell} \cdot \frac{d\ell}{dt} \right]$
 $= u L_l \frac{\partial i}{\partial \ell}.$

Similarly, from (2) we have:- $\frac{\partial i}{\partial \ell} = -C_l \left[-\frac{\partial v}{\partial \ell} \cdot \frac{d\ell}{dt} \right]$
 $= u C_l \frac{\partial v}{\partial \ell}.$

Combining these results, we obtain:-

$$\frac{dv}{di} = \pm \sqrt{\frac{L_l}{C_l}} \quad \text{and} \quad u = \pm \frac{1}{\sqrt{L_l C_l}}.$$

The relative dispositions of voltages, currents, electric and magnetic fields and direction of propagation are illustrated in Fig. 113. The potential difference \vec{V} is in opposition to the electric field \vec{E} , while the current \vec{i} is related to \vec{H} by Ampère's rule. \vec{E} , \vec{H} and \vec{u} form a Right-Handed Set of orthogonal vectors; i.e.,

\vec{E} , \vec{H} and \vec{u} are always arranged at right angles to one another so that a positive (clockwise) rotation through a right angle about \vec{u} will move \vec{E} into the position of \vec{H} . It follows that for two waves travelling in the same direction with \vec{E} vectors alike, the \vec{H} vectors must also be alike, whilst if the \vec{E} vectors are in opposition so are the \vec{H} vectors. A reversal in one but not the other is possible only for a wave travelling in the opposite direction. Of two pulses, travelling in opposite directions along the same transmission line, either the electric or the magnetic fields must be in the same direction, but both cannot be.

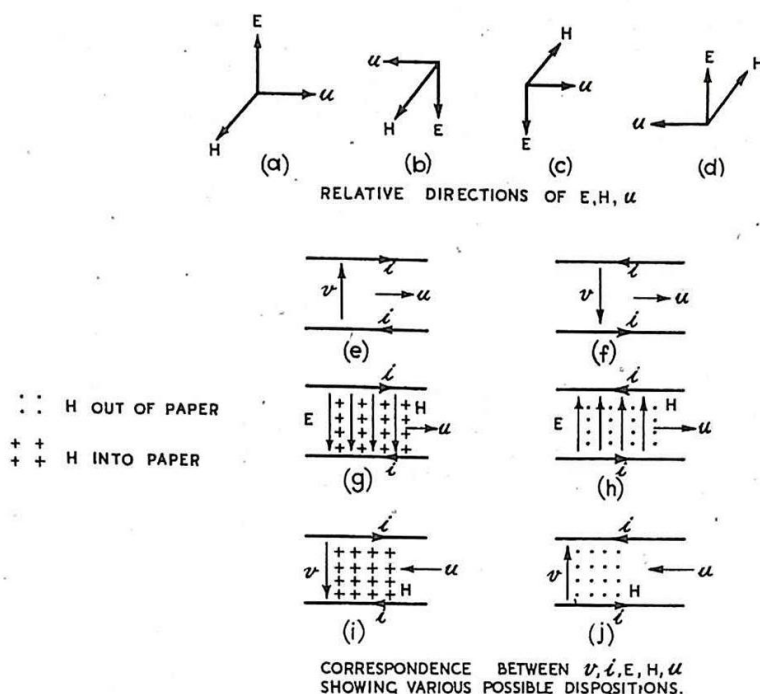


Fig. 113 - Spatial relations between $v, i, E, H,$ and u for a travelling wave.

4. Interference

When two pulses travelling in opposite directions on a line encounter each other, the resultant distribution can be found by adding the voltages and currents of the component waves. It follows from the previous paragraph that wherever such pulses meet there will be a partial cancellation of the magnetic field and an enhancement of the electric, or vice versa. This is illustrated in Fig. 114. Although the resultant wave-pattern is said to be due to the Interference between the two waves, neither affects the other; so called interference is merely a special case of Superposition (see Chap. 1, Sec. 6). This implies that each wave may be considered separate from the other, and the resultant effect obtained by direct addition of currents, fields and voltages.

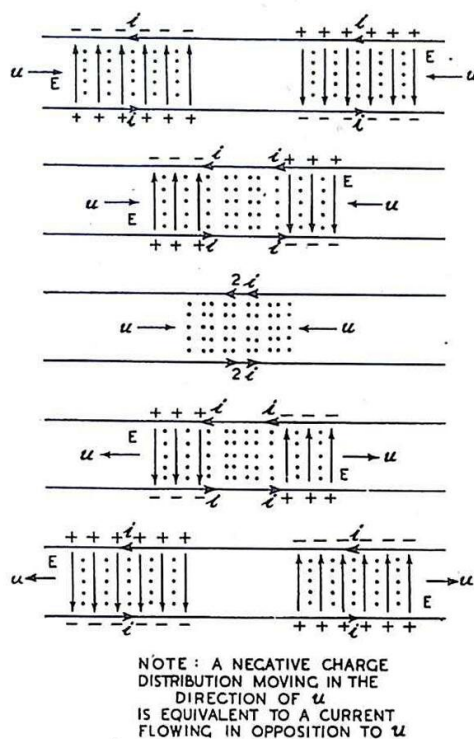


Fig. 114 - Interference between two pulses of equal magnitude.

5. Waves of Various Shapes

A travelling wave on a transmission line may be of any shape, and, if the line is uniform and lossless, it will travel along the line without change of shape, at a uniform velocity independent of the shape of the wave. This follows from the results of Sec. 3 which are independent of the shape of the charge distribution (shown in Fig. 4) and hence the shape of the wave. Figs. 115, 116, and 117 show the line and time-distribution of voltage or current for exponential, rectangular and sinusoidal waves. In each case (a) shows the variation with time of the voltage v_s (or current, since the two are proportional) at the generator or sending end, whilst (b) shows the corresponding instantaneous distribution on the line at the instant $t = t_2$. A construction line shows how one diagram may be obtained from the other; ℓ_1 is the distance travelled in time $(t_2 - t_1)$ by the wave emitted from the generator at any instant t_1 , so that

$$\ell_1 = u(t_2 - t_1).$$

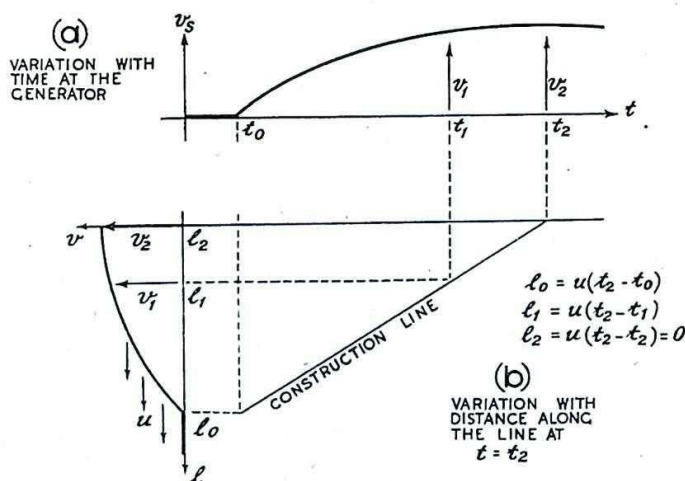


Fig. 115 - Variation of line voltage or current for an exponential wave.

Similar equations hold for the particular cases $\ell_1 = \ell_0$ and $\ell_1 = \ell_2$, corresponding to t_0 and t_2 respectively.

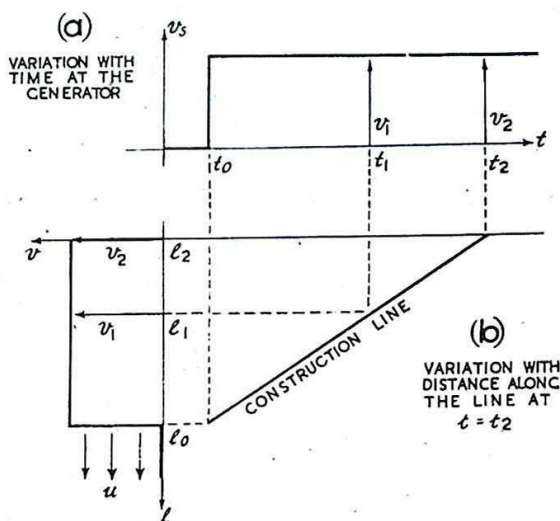


Fig. 116 - Variation of line voltage or current for a rectangular wave.

If the line is an infinite one the impedance which it presents to the generator (Input impedance, Chap. 3) is constant and independent of the type of wave. If finite, but properly terminated, the line appears to the generator to be an infinite line, and behaves as such. If the line is improperly terminated, reflections occur from the end, and interferences between the direct wave and reflected wave causes a change in the input conditions. The nature and extent of these changes will be considered in Secs 8 - 16.

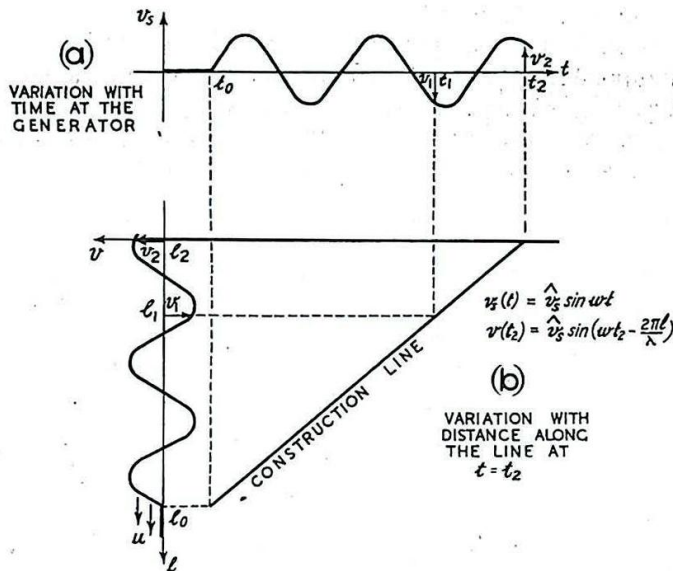


Fig. 117 - Variation of line voltage or current for a sinusoidal wave.

6. Sinusoidal Waves

In the particular case of the sinusoidal wave of Fig. 117, three conditions characteristic of travelling waves are particularly noteworthy:-

(i) The amplitude of the voltage variations is the same for all points of the line, and similarly for the current variations.

This is obvious from preceding paragraphs.

(ii) The variation of phase with distance is given by the relation:-

$$\phi = \phi_s - \frac{2\pi \ell}{\lambda}, \quad \text{for both voltage and current.}$$

ϕ_s is the phase at the sending end.

This follows if we define the wavelength, λ , as the length of the line over which the phase variation at any instant is 2π . Since the phase at the generator varies uniformly with time, and the distance travelled is proportional to time, the phase varies uniformly with the distance along the line.

(iii) The input impedance of the line to an applied sinusoidal EMF is the characteristic impedance $\sqrt{\frac{L}{C}}$.

We have shown that, for a wave travelling along a uniform loss-free line, at any point the ratio of change in voltage to change in current is constant and equal to $\sqrt{\frac{L}{C}}$; this implies that the voltage and current are in phase, i.e. that the impedance is a pure resistance

of value $\sqrt{\frac{L}{C}}$. This impedance is called the Characteristic (or Surge) Impedance of the line, and is denoted by R_0 . (The reciprocal of R_0 , G_0 , is the characteristic admittance. The characteristic impedance is purely resistive only for the case of a distortionless line (see Sec. 19). In the general case of a line which is not loss-free the characteristic impedance and admittance are complex, and are denoted by z_0 and y_0 respectively.)

7. Electromagnetic Waves on Transmission Lines

The types of waves which have so far been considered do not cover all the modes in which electromagnetic energy may be propagated along transmission lines. Such modes may be grouped into two main classes:-

Principal Modes,

Supplementary Modes.

Principal Modes

These are the progressive waves normally used in transmission lines, and those which have been described in Secs. 3 - 6 belong to this category. The notable features of these waves, when the transmission lines are perfectly conducting, are the following:-

- (i) They are propagated at the speed of electromagnetic plane waves in an unbounded medium. This speed u is given by:-

$$u = \frac{c}{\sqrt{K\mu}},$$

where $c = 3 \times 10^8$ m. per second, and K and μ are respectively the dielectric constant and magnetic permeability of the dielectric that fills the space between the conductors. It is independent of frequency.

- (ii) They are transverse electromagnetic waves: that is, the vibrations of E and H take place everywhere in planes (the wave fronts) at right angles to the direction of propagation (along the axes of the lines); also E and H vibrate in phase.
- (iii) The electric field E and magnetic field H are everywhere at right angles in the field pattern.
- (iv) The ratio of the field strengths E and H is the same at all points. Thus in the MKS system of units.

$$\frac{E}{H} = 120\pi\sqrt{\mu/K} \text{ ohms} \quad \begin{matrix} (E \text{ in volts per metre}) \\ (H \text{ in ampères per metre}) \end{matrix}$$

- (v) At a cross-section where the potential difference between the conductors is v a skin current i flows (at high frequencies) in the surface of one conductor and an equal and opposite current in the surface of the other. When a single sinusoidal wave train passes any cross-section the potential v and the current i oscillate in phase and their ratio v/i remains constant and the same at all times for all cross-sections. This ratio which is fixed by the geometry of the system and the dielectric constant of the separating medium is an important property

of the particular transmission line system. It is called the Characteristic Impedance of the system and is usually denoted by the symbol z_0 .

Thus, if v and i are complex numbers representing the voltage and current at any section of the line, then

$$v = z_0 i.$$

Supplementary Modes

Other modes of propagation, each with its own field pattern, can be propagated along a transmission line, in addition to the usual principal mode. In these supplementary modes the electromagnetic field, as in wave guides, possesses either a component of E (E-modes) or of H (H-modes) parallel to the direction of propagation. These modes, like the waves in waveguides, exhibit the phenomenon of cut-off. For each there is a cut-off frequency which is determined by the geometry of the transmission line system. In normal practice the spacing of the transmission lines is small in comparison with the wavelength and the supplementary modes can appear only as Evanescent disturbances, as described in Chap. 5, Sec. 12.

At a geometrical discontinuity in the transmission line system such as a sharp bend or a shunting load, the electromagnetic field assumes a complicated form and in general possesses longitudinal as well as transverse components. It cannot therefore be represented by principal waves alone. Thus, supplementary modes also are excited at the discontinuity. In practice all supplementary modes will be evanescent and at a sufficient distance from the discontinuity their field amplitudes become negligible compared with those of the principal waves.

Evanescent modes carry no power along the transmission lines and their E and H fields are storage fields similar to those of a condenser or an inductance. The storage field excited by a discontinuity does in fact contribute an effective shunting reactance to the transmission line at the discontinuity. A consequence of this reactive behaviour of evanescent modes is that the effective value of the impedance z_l of a circuit component, such as a resistor, when added in shunt across the line may be quite different from its nominal value z_n that would be anticipated from its electrical behaviour at low frequencies.

The presence of evanescent modes begins to be important at wavelengths of $1\frac{1}{2}$ metres and is very important at wavelengths of 10 centimetres and less. At wavelengths greater than a few metres the effect of the storage fields of the evanescent modes excited at discontinuities may be neglected.

The usual theory of transmission lines is the theory of the principal wave and ignores the behaviour of supplementary modes. From what has preceded it is apparent why the standard treatment is adequate when the wavelength exceeds a few metres but gives an incomplete description of transmission line phenomena at wavelengths less than $1\frac{1}{2}$ metres.

UNIFORM LOSS-FREE LINES: REFLECTION AT THE TERMINATION, AND ITS EFFECTS

8. Reflection of a Rectangular Wave

We shall assume that the line has a characteristic impedance which is purely resistive and of value R_0 , independent of frequency, and that the electromagnetic disturbances considered are propagated without distortion along the line with uniform velocity u .

A DC generator whose output resistance is R_0 , is connected to a section of uniform line of length ℓ , terminated in R_r , which is not necessarily equal to R_0 . The circuit is shown in Fig. 118(a). Initially the line is uncharged, and the current everywhere is zero. As the switch S is closed, a wave begins to travel from the sending to the receiving end of the line. Current flows in both conductors as illustrated in Fig. 118(b) and the line becomes progressively charged.

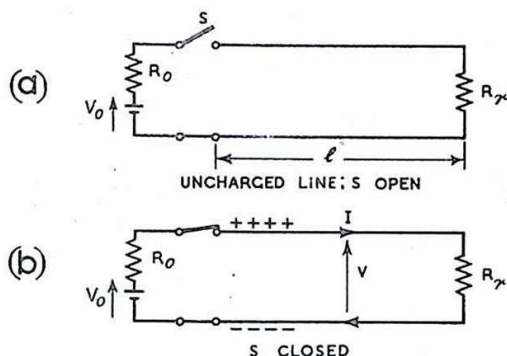


Fig. 118 - Application of a rectangular wave to a transmission line,

If the open-circuit EMF of the generator is V_0 volts, the voltage V_+ which is impressed on the line is $\frac{V_0}{2}$, and the current I_+ which flows in the travelling wave is

$$\frac{V_+}{R_0} = \frac{V_0}{2R_0}.$$

[The suffixes + and - refer to the waves travelling away from and towards the generator respectively. The suffixes r and s refer to the receiving and sending end respectively.]

When the wave front reaches the receiving end current begins to flow through the load, and a potential difference is set up across the load resistance R_r . If $R_r = R_0$, the energy of the wave is completely absorbed in the termination, and the same conditions obtain on the line as if it were part of an infinite line. The energy absorbed by the load must be supplied by a load voltage V_ℓ and current I_ℓ satisfying the relation

$$\frac{V_\ell}{I_\ell} = R_r,$$

and if $\frac{V_+}{I_+}$, which equals R_0 , is not equal to R_r , such conditions cannot be maintained without some energy from the direct wave being "rejected" by the load. A wave is then reflected back towards the sending end as if the termination behaved partly as an absorber of energy and partly as a new generator. If $R_r > R_0$, the voltage developed across the load is greater than V_+ , whilst if $R_r < R_0$, the current through the load is greater than I_+ . In both cases the resultant load voltage and current must arrange themselves in the ratio

$$\frac{V_\ell}{I_\ell} = R_r.$$

It is convenient to represent the line as consisting of an upper and a lower conductor having a sending end on the left and a receiving end on the right, and to adopt the following sign conventions:-

- (i) for a wave travelling away from the sending end (direct wave)
voltage is measured positively from the lower to the

upper conductor, conventional current is measured positively to the right in the upper conductor, and to the left in the lower conductor.

- (ii) for a wave travelling towards the sending end (reflected wave)

voltage is measured as before, positively from the lower to the upper conductor; conventional current is measured positively to the left in the upper and to the right in the lower conductor.

9. Coefficient of Reflection

Suppose a fraction ρ of the sending wave of Sec. 8 is reflected, so that the resultant load voltage is $V_+ + \rho V_+$ (the charge density of the reflected wave being ρ times that of the direct wave). Since the ratio of voltage to current in any wave transmitted by the line is R_0 , the magnitude of the current in the reflected wave is $\rho \frac{V_+}{R_0} = \rho I_+$.

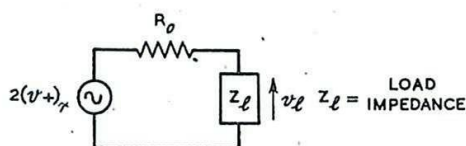


Fig. 119 - Equivalent circuit.

Thus $V_L = V_+ + \rho V_+$ and $I_L = I_+ - \rho I_+$, the current of the reflected wave being in opposition to that of the direct wave.

$$\text{Hence } \frac{V_L}{I_L} = \frac{V_+ (1 + \rho)}{I_+ (1 - \rho)}$$

$$\text{i.e. } R_T = R_0 \frac{(1 + \rho)}{(1 - \rho)}$$

$$\text{i.e. } \rho = \frac{R_T - R_0}{R_T + R_0}$$

ρ is called the Coefficient of Reflection. Its magnitude varies between +1 (for $R_T = \infty$) and -1 (for $R_T = 0$); and it is zero when $R_T = R_0$. (The case where ρ is complex is dealt with briefly in Sec. 13).

It is sometimes convenient to use the equivalent circuit shown in Fig. 119 for deriving the voltage and current produced at the load by a wave which travels to, and is partly reflected from, the termination. The equivalent generator voltage $2(v_+)_r$ is twice the instantaneous value of the voltage of the direct wave measured at the receiving end. This circuit is valid for any type of wave and is particularly useful in transient problems where the nature of the direct wave is known, since it gives the load voltage and current directly without the resultant line voltage having to be determined first.

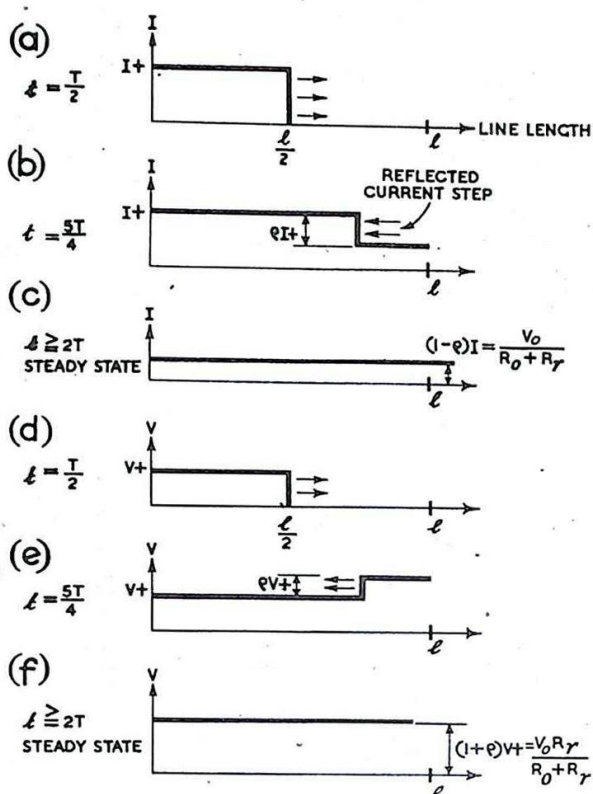


Fig. 120 - Current and voltage distribution on the line of fig. 118 at various intervals after closure of switch S (Drawn for $R_r > R_0$).

10. Line Fed by a DC Generator

The voltage and current distribution on the line of Fig. 118 at various intervals after the closure of the switch S are shown in Fig. 120. The conditions at the input terminals are shown plotted against time in Figs. 121 and 122. It should be noted that since the generator output impedance is R_0 , the conditions at the load after time T , and at the generator after time $2T$ are the same as if the line were a "short-circuit", in series with R_r . If the generator output impedance were other than R_0 , further diminishing reflections would occur alternately at both ends and would persist until the steady condition were reached corresponding, as before, to the solution which would be obtained by simple direct-current-theory.

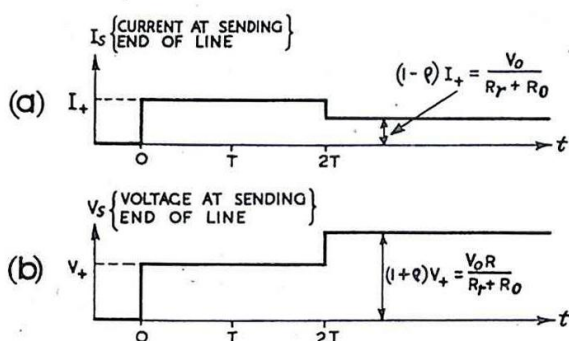


Fig. 121 - Conditions at input terminals of line shown in fig. 118, drawn for $R_r > R_0$ (ρ is positive).

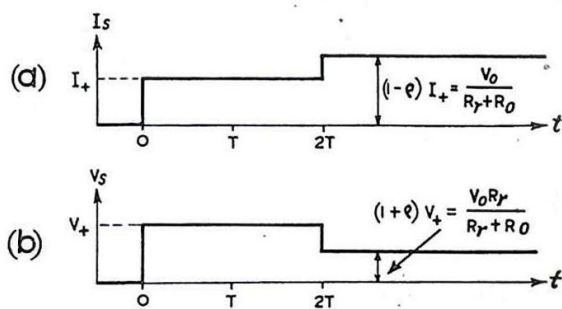


Fig. 122 - Conditions at input terminals of line shown in fig. 118, drawn for $R_r < R_0$ (ρ is negative).

For practical purposes, except in cases of severe mismatching, the steady state is reached to a sufficient degree of approximation after a few transits of the line.

11. Deduction of Line Characteristics

The fundamental characteristic properties of a uniform lossless transmission line have been deduced in Sec. 3. An alternative derivation, based on energy considerations, is given below. We assume certain properties already demonstrated, namely, that the line is distortionless and has a characteristic impedance which is a pure resistance; also that the wave of Figs. 118 - 122 takes a finite time to travel along the line, so that the conditions at the sending end remain constant from the closure of the switch until at least $t = 2T$, when the front of the reflected wave, if any, returns. Thus, whether the line is short-circuited or open-circuited, or has any other termination, cannot be known at the generator until $t = 2T$.

If the total line capacitance is C , and the total inductance L , the following relationships obtain for a generator of internal resistance R_0 .

At the end of time $t = 2T$ the conditions are:-

Short-circuited line

$I_s = 2I_+$, $V_s = 0$, so that there is no energy stored in the electric field. The energy stored in the magnetic field is

$$\begin{aligned} & \frac{1}{2} L (2I_+)^2 \\ &= 2 L I_+^2 \end{aligned}$$

Open-circuited line

$V_s = 2V_+$, $I_s = 0$, so that there is no energy stored in the magnetic field. The energy stored in the electric field is

$$\begin{aligned} & \frac{1}{2} C (2V_+)^2 \\ &= 2 C V_+^2 \end{aligned}$$

In both cases the energy supplied by the generator is $V_+ I_+ \cdot 2T$.

$$\text{Hence} \quad 2 V_+ I_+ T = 2 L I_+^2 = 2 C V_+^2$$

$$\text{i.e.} \quad \frac{V_+}{I_+} = \frac{L}{T} = \frac{T}{C}.$$

Therefore $T^2 = LC$ so that $T = \sqrt{LC}$;

$$\text{and } R_0 = \frac{V_+}{I_+} = \sqrt{\frac{L}{C}}.$$

If the length of the line is ℓ metres:-

$\frac{L}{\ell} = L_\ell$ is the inductance per unit length;

$\frac{C}{\ell} = C_\ell$ is the capacitance per unit length.

$$\begin{aligned} \text{Then } T &= \sqrt{LC} = \sqrt{Ll \cdot Cl} \\ &= l \sqrt{Ll Cl} \end{aligned}$$

Therefore the velocity of propagation is $u = \frac{l}{T}$

$$= \frac{1}{\sqrt{Ll Cl}}$$

$$\text{Also } R_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{Ll}{Cl}} = \sqrt{\frac{Ll}{Cl}}$$

12. Discharge of Open-circuited Line

In the arrangement shown in Fig. 123 a section of uniform loss-free cable has been charged, so that there is a potential difference of V_0 volts between the conductors, which are not connected at either end. The switch S is closed at $t = 0$ and the line begins to discharge through the load resistance R_l .

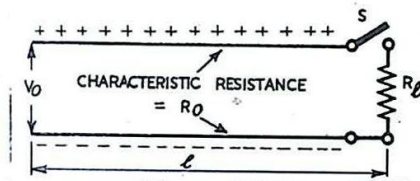


Fig. 123 - Discharge of open-circuited line.

The drop of voltage ($V_0 - V_1$) at the end which is suddenly terminated in R_l may be regarded as a rectangular voltage wave travelling away from R_l , as shown in Fig. 124 (a) and (b).

When this wave reaches the open circuit after time T , the entire line has been discharged to a new level V_1 . At this instant total reflection occurs and a second voltage wave, also of magnitude ($V_0 - V_1$), begins to travel from the open circuit towards R_l (Fig. 124(b) and (c)). This wave will reach R_l in time T after its inception, i.e. in time $2T$ from the closure of the switch. Prior to this time the load R_l must remain insensitive to any reflection effects.

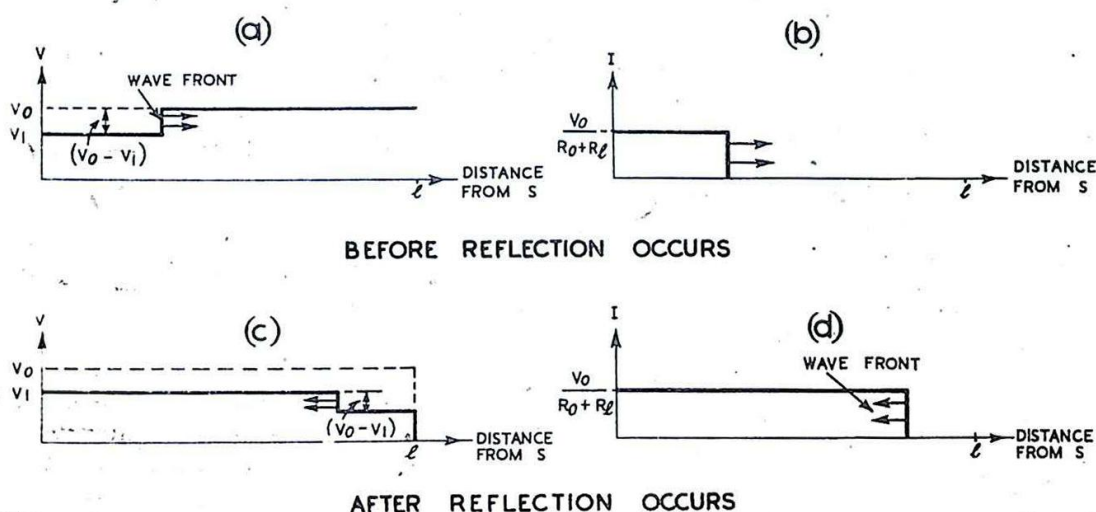


Fig. 124 - Line distribution of voltage and current before and after reflection occurs.

The voltage developed across the load is initially V_1 ; it remains at this level for an interval $2T$, and then changes to V_2 . We shall consider the relative magnitudes of V_0 , V_1 , V_2 etc., in relation to R_0 and R_l .

Initially, the line acts a generator of EMF V_0 and output resistance R_0 ; thus the voltage developed across the load is given by:-

$$V_1 = V_0 \frac{R_l}{R_0 + R_l}.$$

A negative voltage wave of magnitude $V_0 \frac{R_0}{R_0 + R_l}$ travels along the line towards the open circuit. This is reflected from the open circuit, leaving the line voltage

$$V_0 - \frac{2 V_0 R_0}{R_0 + R_l} = V_0 \frac{(R_l - R_0)}{R_l + R_0}$$

$= \rho V_0$, where ρ is the reflection coefficient for the load R_l . The line current is then zero. Immediately after the first period $2T$ the conditions are the same as at the closure of the switch, except that the line is charged to a voltage ρV_0 instead of V_0 . The whole process is repeated, with $V_2 = \rho V_1$, and after time $4T$ the line voltage is reduced everywhere to $\rho^2 V_0$ and a third voltage step $\rho^2 V_1$ begins to appear across the load.

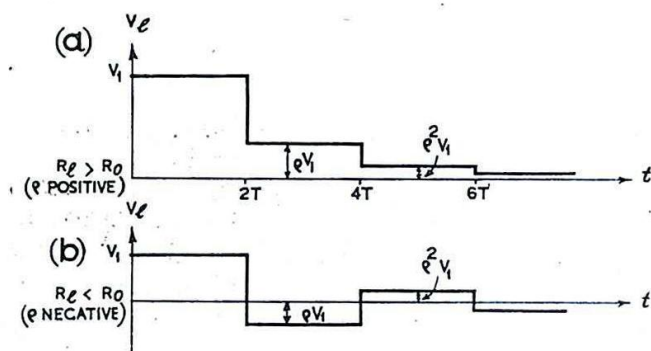
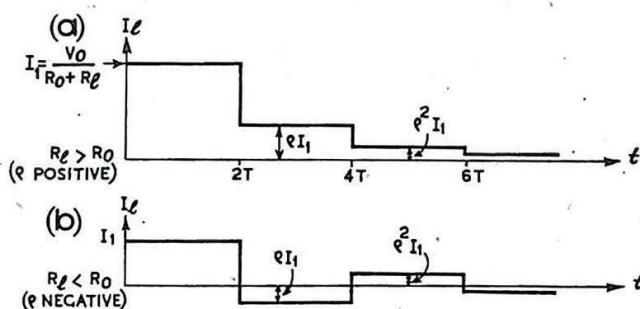


Fig. 125 - Voltage V_l developed across load with (a) ρ positive (b) ρ negative.

Fig. 126 - Current I through load with (a) ρ positive and (b) ρ negative.



The sequence is pictured in Figs. 125 and 126. In each figure (a) refers to the case where $R_l > R_0$ and (b) refers to the case where $R_l < R_0$. The decay in the energy stored in the line (Fig. 127) is exponential, similar to the decay of energy in a discharging condenser, except that the process is stepped in the former case and smooth in the latter. If $\rho = 0$ i.e. if $R_l = R_0$, the line is completely discharged after $2T$ secs.

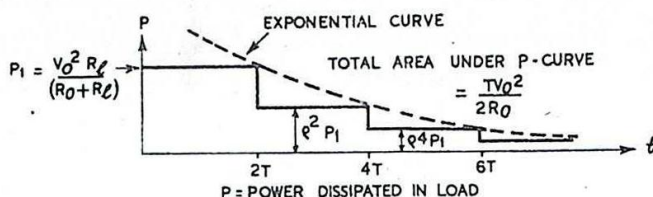


Fig. 127 - Power dissipated in load.

13. Reflection of Sine Waves

The instantaneous conditions of reflection at the receiving end of the line are exactly the same for a sinusoidal as for a constant voltage wave, but the pattern of the interference between the direct and reflected waves is more complicated because of the sinusoidal variations. These two sine waves, travelling in opposite directions, interact as illustrated in Fig. 128.

(The suffix + is used to indicate the direct, and -, the reflected wave.) Thus v and i , the resultant voltage and current at any point on the line are given by:-

$$v = v_+ + v_-, \quad i = i_+ - i_-$$

$$\text{At the termination, } (v_-)_r = \rho (v_+)_r$$

$$\text{and } (i_-)_r = \rho (i_+)_r$$

The phase of v_+ at a distance l from the termination, is advanced on that of $(v_+)_r$ by $\frac{2\pi l}{\lambda}$ radians; whilst the phase of v_- ,

at the same point, is delayed on that of $(v_-)_r$ by the same amount.

(It is commonly found that these statements are not readily appreciated. What should be clear is that the wave which has travelled farther must be more "stale", since it issued from the generator earlier. The later it issues from the generator, the more recent the phase.)

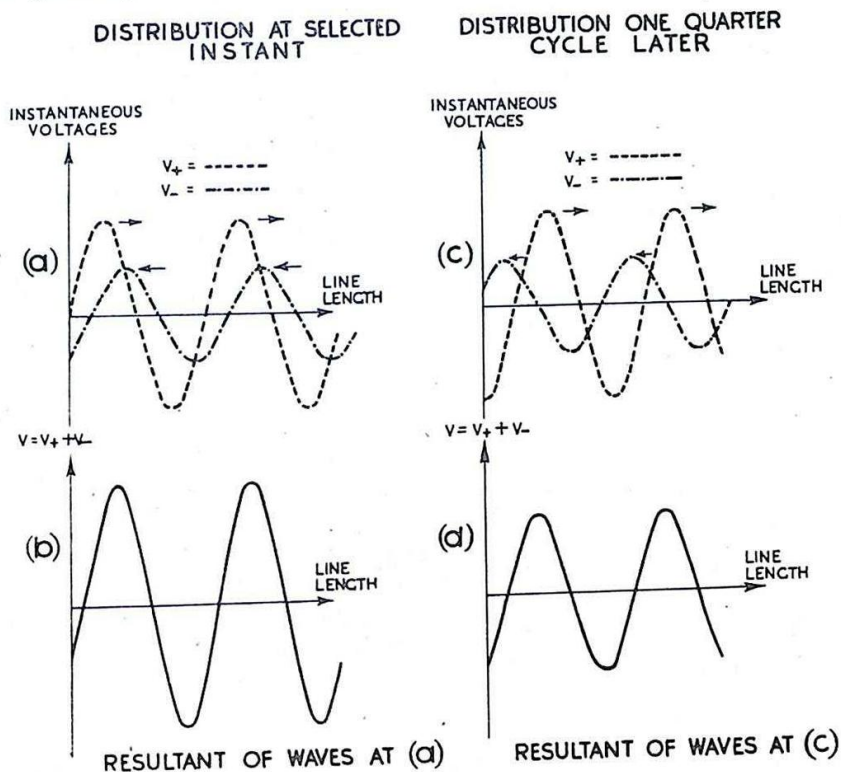


Fig. 128 - Interference between direct and reflected sinusoidal waves.

If the terminating impedance is not a pure resistance there will be a phase difference between $(v_+)_r$ and $(v_-)_r$; in other words, ρ will be complex. In this case

$$\rho = \frac{(v_-)_r}{(v_+)_r} \quad \text{and} \quad |\rho| = \frac{(\hat{v}_-)_r}{(\hat{v}_+)_r} = \frac{\hat{v}_-}{\hat{v}_+},$$

since the amplitudes of the direct and reflected waves are the same at all other points as at the termination, the line being loss-free.

The phase difference between v_+ and v_- will in any case vary with l , the distance from the receiving end, and will increase by 2π as l increases by $\frac{\lambda}{2}$. As l is varied there will be found

points at $\frac{\lambda}{2}$ intervals at which v_+ and v_- are always anti-phase, and others midway between them at which they are always in phase. The former points are called voltage nodes, and the peak value of the voltage at these points is $\hat{v}_+ - \hat{v}_- = \hat{v}_+ (1 - |\rho|)$.

The intermediate points are called voltage antinodes and the peak voltage there is $\hat{v}_+ + \hat{v}_- = \hat{v}_+ (1 + |\rho|)$.

It may be shown that voltage nodes are current antinodes, and vice versa. At a current node, the peak value is $\hat{i}_+ (1 - |\rho|)$ and at an antinode, $\hat{i}_+ (1 + |\rho|)$.

The ratio $\frac{\text{maximum peak value}}{\text{minimum peak value}}$ for either current or voltage is called the Standing Wave Ratio (SWR). For a loss-free line, the SWR is given by:-

$$S = \frac{\frac{\hat{v}_+}{\hat{v}_+ - \hat{v}_-}}{\frac{\hat{v}_+}{\hat{v}_+ + \hat{v}_-}} = \frac{1 + |\rho|}{1 - |\rho|} = \frac{\hat{i}_+ + \hat{i}_-}{\hat{i}_+ - \hat{i}_-}.$$

This definition will be generalised later to include lossy lines, where there is a SWR for each point of the line (Sec. 24).

At a voltage antinode,

$$\hat{v} = \hat{v}_+ (1 + |\rho|) \quad \text{and} \quad \hat{i} = \hat{i}_+ (1 - |\rho|).$$

At such a point the magnitude of the impedance is a maximum given by:-

$$\hat{Z} = \frac{\hat{v}_+}{\hat{i}_+} \left(\frac{1 + |\rho|}{1 - |\rho|} \right),$$

$$= R_0 S.$$

Hence
$$S = \frac{\hat{Z}}{R_0}.$$

Similarly it may be shown that, at a voltage node, the magnitude of the impedance is a minimum \hat{Z} and that $S = \frac{R_0}{\hat{Z}}$.

At all nodes and antinodes the impedance is purely resistive. Fig. 128 shows how, in the region of interference, the three properties which were found to hold for travelling waves (Sec. 6), are changed fundamentally to become:-

- (i) The amplitude of both voltage and current varies from point to point on the line.
- (ii) The variation of phase with distances is, in general, much more complicated than for travelling waves.
- (iii) The input impedance of the line depends on both the line length and the termination.

It is useful to consider the particular cases of open-circuited and short-circuited lines. It should be noted that the direct wave may always be divided into two parts, one given by $(1 - \rho)v_+$ and the other by ρv_+ . The second part is identical with the reflected wave except in direction of propagation, and the interaction between them is the same as that which follows reflection at an open circuit. Hence the composite wave may be split into a travelling wave $(1 - \rho)v_+$ and what is called a Standing Wave, of amplitude $2|\rho|\hat{v}_+$, as will be shown in subsequent sections. Energy is transmitted by the travelling wave only, the standing wave storing, and not transmitting, energy.

When the amplitudes of reflected and direct waves are equal ($\rho = \pm 1$) the standing wave produced is sometimes called a Complete Standing Wave. In such a case the voltage and current falls to zero at the respective nodes. Otherwise the distribution may be termed a Partial Standing Wave.

14. Standing Waves on a Uniform Open-Circuited Loss-Free Line

The line distribution of the steady voltage and current at successive intervals is shown in Fig. 129 for a uniform loss-less line terminated in an open circuit. Direct and reflected waves are shown, together with the resultant voltage and current distributions. It will be seen that at even multiples of $\lambda/4$, (multiples of $\frac{\lambda}{2}$) from the termination, the amplitude of the current is zero

and that of the voltage is $2\hat{v}_+$. Similarly at odd multiples of $\frac{\lambda}{4}$

from the termination, the voltage is always zero and the maximum value of the current is $2\hat{i}_+$. The voltages at all points between two consecutive voltage nodes are in phase, and are antiphase with those of the adjoining half-wavelength sections. At any point the current and voltage are in quadrature.

The relation between voltage and current at any point on the line may be determined from the vector diagrams of Fig. 130. The phase of $(v_+)_r$ is taken as the standard phase, and this lags the phase of v_+ , a distance l from the open circuit, by a phase angle ϕ radians where $\phi = \frac{2\pi l}{\lambda}$. Similarly, $(v_-)_r$ leads v_- by the same phase angle.

For $0 < l < \frac{\lambda}{4}$, the phase of the resultant current is shown in the figure (b) to lead that of the voltage in quadrature, i.e., the short section behaves, in the steady state, like a pure capacitance, C_0 , (the equivalent capacitance).

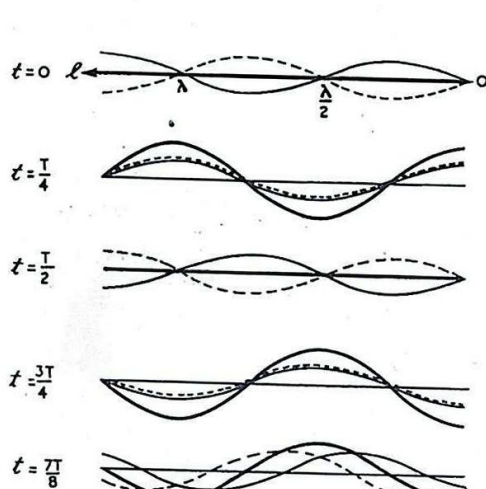
$$\text{where } \frac{1}{\omega C_0} = -X = R_0 \cot \phi.$$

For $\frac{\lambda}{4} < l < \frac{\lambda}{2}$, v leads i in quadrature (c), so that the line

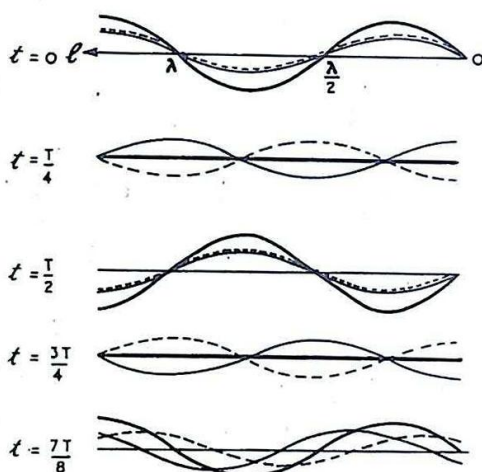
"looks like" a pure inductance L_e where $\omega L_e = X = -R_0 \cot \phi = R_0 \tan(\phi - \frac{\pi}{2})$.

For larger values of l the effective capacitance and inductance repeat for every additional quarter-wavelength of the line.

The input reactance is given for all values of l by the equation $X = -R_0 \cot \phi$. Its variation with l is shown in Fig. 131.



VOLTAGE WAVES ON AN OPEN-CIRCUITED LINE



CURRENT WAVES ON AN OPEN-CIRCUITED LINE

— DIRECT TRAVELLING WAVE
 --- REFLECTED TRAVELLING WAVE
 ——— RESULTANT STANDING WAVE

l = DISTANCE FROM TERMINATION 0

Fig. 129 - Line distribution of voltage and current at successive intervals for uniform open-circuited loss-free line (steady state).

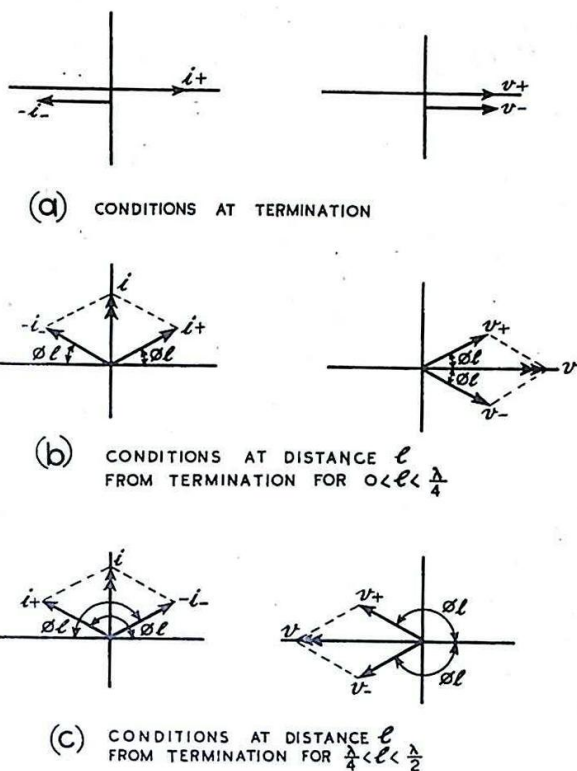


Fig. 130 - Vector diagrams showing relation between voltage and current at various distances from the termination (open-circuited line).

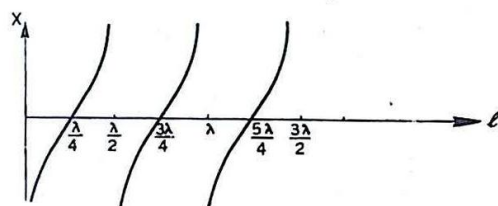
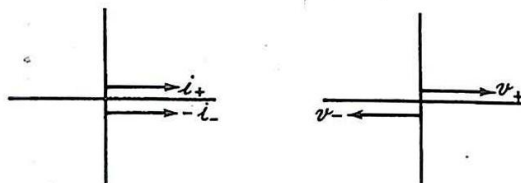
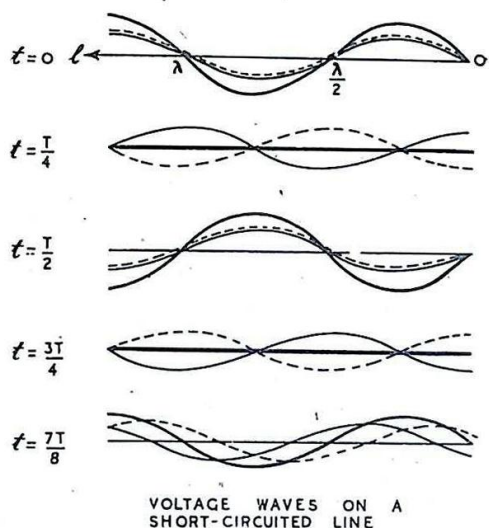


Fig. 131 - Variation of input reactance of open-circuited line with line-length.

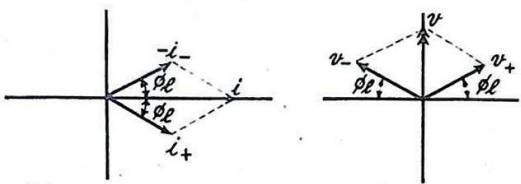
15. Standing Waves on a Uniform Short-Circuited Loss-Free Line

The conditions maintained at a short circuit, viz: $v_r = 0$, $\rho = -1$, are the same as those which arise on an open-circuited line at a voltage node, i.e., at odd multiples of $\frac{\lambda}{4}$ from the open circuit.

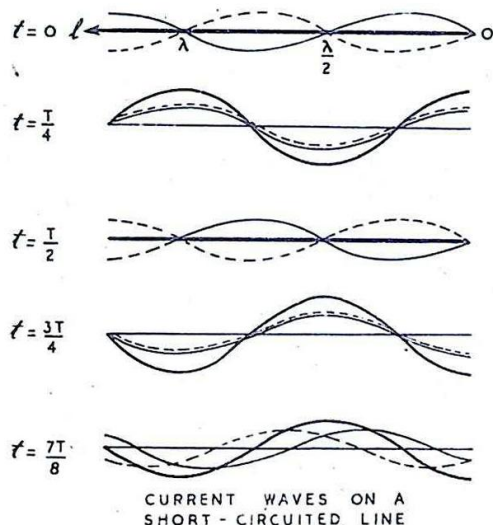
Since the standing wave distribution for the open-circuit case has already been established, it is convenient to deduce that for the short-circuit case by omitting from the former diagrams the last $\frac{\lambda}{4}$ section of the line.



(d) CONDITIONS AT TERMINATION

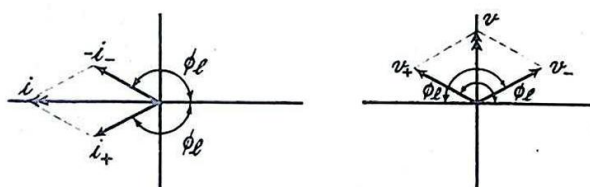


(b) CONDITIONS AT DISTANCE l FROM TERMINATION FOR $0 < l < \frac{\lambda}{4}$



— DIRECT TRAVELLING WAVE
 --- REFLECTED TRAVELLING WAVE
 — RESULTANT STANDING WAVE

Fig. 132 - Line distribution of voltage and current at successive intervals for uniform short-circuited loss-free line (steady state).



(c) CONDITIONS AT DISTANCE l FROM TERMINATION FOR $\frac{\lambda}{4} < l < \frac{\lambda}{2}$

Fig. 133 - Vector diagrams showing relation between voltage and current at various distances from the termination (short-circuited line).

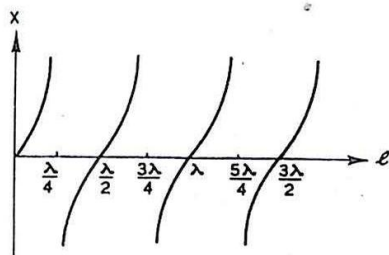


Fig. 134 - Variation of input reactance of a short-circuited line with line-length.

The line distribution thus obtained is shown in Fig. 132, and Fig. 133 gives the vector diagrams showing the relation between voltage and current at the sending end of various lengths of line in the steady state. Fig. 134 gives the variation of input reactance with line length.

Analytically, it will be seen that the two sets of diagrams for the open-circuit and short-circuit cases are mutually interchangeable by substituting voltage in one for current in the other, and vice versa.

It follows that the input susceptance of a short-circuited line of length l is $B = \frac{-1}{R_0} \cot \phi$, so that the input reactance is $X = R_0 \tan \phi$, as shown in Fig. 134.

16. Input Impedance of a Uniform Loss-Free Line for any Termination

For a line of length l and characteristic impedance R_0 , terminated in z_r , it may be shown that the input impedance is:-

$$z_s = R_0 \cdot \frac{z_r + jR_0 \tan \frac{2\pi l}{\lambda}}{R_0 + j z_r \tan \frac{2\pi l}{\lambda}}$$

If z_r is a pure reactance, so is z_s .

If $z_r = R_0$, $z_s = R_0$, as has already been shown.

If z_r is a pure resistance other than R_0 , or is partly reactive, then z_s is in general neither a pure resistance nor a pure reactance.

If z_r is a pure resistance equal to R_r , the input impedance is resistive at all multiples of $\frac{\lambda}{4}$ from the receiving end; the corresponding values of z_s are R_r at multiples of $\frac{\lambda}{2}$ from the termination, and $\frac{R_0^2}{R_r}$ at odd multiples of $\frac{\lambda}{4}$. For all other lengths of line the magnitude of the impedance lies between these two limits, R_r and $\frac{R_0^2}{R_r}$.

On a loss-free line the input impedance is always resistive at nodes and antinodes, and alternate values have a geometric mean R_0 . Also, at intermediate $\lambda/4$ points the magnitude of the impedance is R_0 .

FUNDAMENTAL LINE CHARACTERISTICS

17. Distributed Elements for a Loss-Free Line

We have shown in Secs. 3 to 6 that if a line is capable of transmitting an electromagnetic wave without loss, it presents a resistive impedance to a generator connected to it. The velocity of propagation is

$$u = \frac{1}{\sqrt{L_l C_l}} \text{ and the value}$$

of the characteristic resistance is $\sqrt{\frac{L_l}{C_l}}$ where L_l and C_l

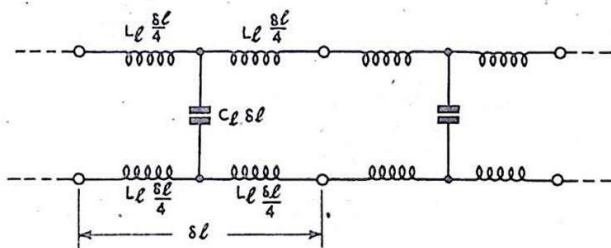


Fig. 135 - Representation of uniform lossless transmission line as an equivalent ladder of infinitesimal filter sections.

C_ℓ are the values of the inductance and capacitance, per unit length, respectively. These results, when compared with those of Chap. 3, Sec. 10, suggest that a line may be represented by an equivalent arrangement of low-pass filter sections in cascade, each section being infinitesimally small. This conception is illustrated in Fig. 135. Each section represents a length $\delta\ell$ of the line. Using the results of Chapter 3, Section 10, we have:-

$$f_c = \frac{1}{\pi \delta\ell \sqrt{L_\ell C_\ell}}, \text{ and } z = \sqrt{\frac{L_\ell}{C_\ell}} \cos \frac{\delta\beta}{2},$$

$$\text{where } \sin \frac{\delta\beta}{2} = \frac{f}{f_c}; \quad \delta\beta \text{ being the phase shift for}$$

the section.

$$\text{As } \delta\ell \rightarrow 0, f_c \rightarrow \infty, \delta\beta \rightarrow 0, \sin \frac{\delta\beta}{2} \rightarrow 0 \text{ and } \cos \frac{\delta\beta}{2} \rightarrow 1.$$

From these results the characteristic properties of transmission lines may be obtained as follows.

Characteristic impedance

$$z_0 = \sqrt{\frac{L_\ell}{C_\ell}}, \text{ a pure resistance, } (R_0).$$

Attenuation

Since $f_c \rightarrow \infty$ as $\delta\ell \rightarrow 0$, the cut-off frequency is infinite and there is no attenuation, but signals at all frequencies are transmitted equally well.

Phase distribution and velocity

Assuming that the phase delay $\delta\beta$ is due to a time delay δt , we may determine $\frac{\partial\beta}{\partial\ell}$ and $\frac{d\ell}{dt}$; i.e., the variation of phase with

distance along the line at any instant, and the velocity u with which any point of the wave, at which the phase is constant, moves along the line.

From the above relations, it follows that

$$\frac{\partial\beta}{\partial\ell} \doteq -2\pi f \sqrt{L_\ell C_\ell} = -\omega \sqrt{L_\ell C_\ell}; \text{ the negative sign indic-}$$

ates that there is a lag in phase as the distance from the generator is increased.

It may be shown that, in the limit, as $\delta\ell \rightarrow 0$,

$$\frac{\partial\beta}{\partial\ell} = -\omega \sqrt{L_\ell C_\ell};$$

i.e., at any instant the change of phase along a given length of the line is constant and proportional to the length.

In particular, if the change in β is 2π radians the line length over which this occurs is defined as λ , the wavelength, and is given by:-

$$\lambda = \frac{2\pi}{\omega \sqrt{L_\ell C_\ell}} = \frac{1}{f \sqrt{L_\ell C_\ell}}.$$

u, the phase velocity is given by:-

$$u = \frac{d\ell}{dt} = - \frac{\frac{\partial \ell}{\partial \beta}}{\frac{\partial \ell}{\partial \omega}} = \frac{-\frac{\partial \beta}{\partial \omega}}{\frac{\partial \beta}{\partial \omega}} = - \frac{\omega}{-\omega \sqrt{L_\ell C_\ell}} = \frac{1}{\sqrt{L_\ell C_\ell}} .$$

It follows that $\lambda = \frac{u}{f}$; i.e. λ is the distance travelled by the wavefront in the time occupied by one cycle at the generator.

Propagation constant

Since there is no change in amplitude but a phase delay $\beta_\ell = - \frac{\partial \beta}{\partial \ell} = \frac{2\pi}{\lambda}$ per unit length, the propagation constant γ for a length ℓ of the line may be written

$$\gamma = j\beta = j\beta_\ell \ell = j \frac{2\pi \ell}{\lambda} .$$

If $v_s = \hat{v}_s e^{j\omega t}$ at the generator, the voltage at a point distant ℓ from the generator is given by:-

$$v = v_s e^{-j\frac{2\pi \ell}{\lambda}} = \hat{v}_s e^{j(\omega t - \frac{2\pi \ell}{\lambda})} .$$

18. Distributed Elements for a Lossy Line

In general it is impossible to ignore the effect of resistance of the conductors and leakage through the dielectric, and these modify the line characteristics.

Fig. 136 shows the modifications necessary to introduce these quantities into the infinitesimal section of Fig. 135.

Analysis of such a section shows that, on proceeding to the limit as $\delta \ell \rightarrow 0$, the characteristic impedance z_0 , and the propagation constant γ_ℓ for a line of length ℓ , are given by:-

$$z_0 = \sqrt{\frac{R_\ell + j\omega L_\ell}{G_\ell + j\omega C_\ell}} ; \quad \gamma = \ell \sqrt{(R_\ell + j\omega L_\ell)(G_\ell + j\omega C_\ell)} .$$

We may write $\gamma = \gamma_\ell L$, where γ_ℓ is the propagation constant per unit length; whence we have that

$$\gamma_\ell = \sqrt{(R_\ell + j\omega L_\ell)(G_\ell + j\omega C_\ell)}$$

$$= \alpha_\ell + j\beta_\ell$$

(in nepers) per unit length, and β_ℓ the phase shift (in radians) per unit length.

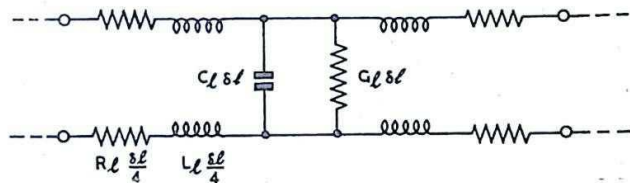


Fig. 136 - Representation of uniform lossy transmission line as an equivalent ladder of infinitesimal filter sections.

Since α_l is not in general constant, and β_l is not proportional to frequency, amplitude and phase distortion occur; and unless sinusoidal, the shape of a wave changes as it travels along the line.

Since α_l is not zero, there is attenuation as well as phase shift, energy being dissipated in the conductors (because of R_l) and in the dielectric (because of G_l) as the wave progresses.

19. Characteristic Impedance

$$z_0 = \sqrt{\frac{R_l + j\omega L_l}{G_l + j\omega C_l}}.$$

In general, this is complex. In the particular case, where $\frac{L_l}{C_l} = \frac{R_l}{G_l}$, the line is distortionless and z_0 is a pure resistance of magnitude $\sqrt{\frac{L_l}{C_l}}$; but this is not a case commonly encountered in radar. Since line losses are wasteful, every effort is made to reduce them to a minimum. In particular, matching devices become inefficient unless constructed of lines with very low losses.

It may be shown that for open-wire feeders in which conductors each of radius r are spaced with their centres a distance d apart, the characteristic impedance is:-

$$R_0 \doteq 276 \cdot \log_{10} \frac{d}{r} \dots \dots \dots (1).$$

The corresponding formula for low-loss coaxial cables is

$$R_0 = 138 \log_{10} \frac{r_2}{r_1} \dots \dots \dots (2),$$

where r_2 is the internal radius of the outer, and r_1 the radius of the inner conductor.

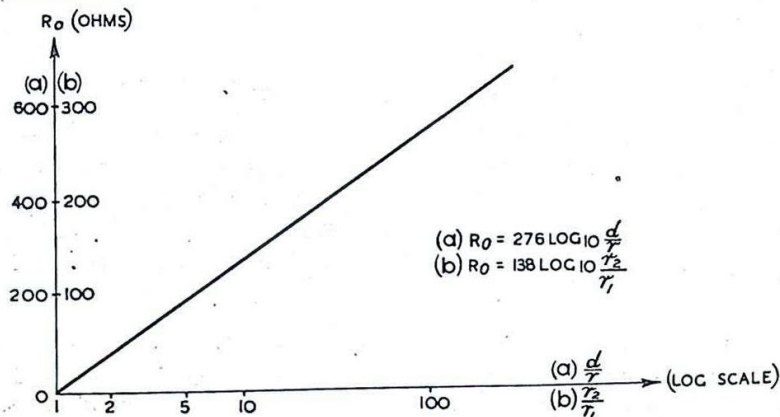


Fig. 137 - Variation of R_0 with frequency for (a) open wire lines, ($d \gg r$), (b) coaxial cables.

These formulae neglect losses, and the former is an approximation which assumes that $d \gg r$. They are illustrated in Fig. 137.

If in the open-wire feeder r is comparable with d the exact formula $R_0 = 120 \cosh^{-1}(d/2r)$ must be used. This is illustrated in Fig. 138.

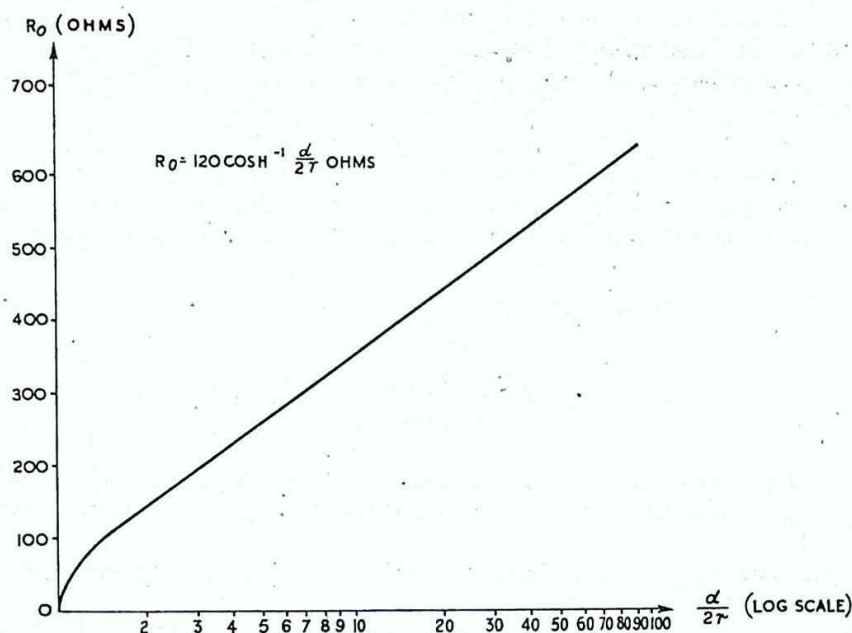


Fig. 138 - Exact relation between R_0 and d/r for open-wire lines formed of twin cylinders.

For conductors embedded in a non-magnetic material of dielectric constant K , the formulae become:-

$$R_0 \doteq \frac{276}{\sqrt{K}} \log_{10} \frac{d}{r} \dots\dots\dots(1a)$$

and $R_0 \doteq \frac{138}{\sqrt{K}} \log_{10} \frac{r_2}{r_1} \dots\dots\dots(2a)$

20. Propagation Constant

$$\gamma_l = \sqrt{(R_l + j\omega L_l)(G_l + j\omega C_l)} = \alpha_l + j\beta_l$$

In the particular case of the distortionless line, mentioned in Sec. 19, $\alpha_l = \sqrt{R_l G_l}$ and $\beta_l = \omega \sqrt{L_l C_l}$, and these satisfy the conditions for distortionless transmission; viz: attenuation independent of frequency and phase shift proportional to frequency. This ensures that for a wave of any shape the sinusoidal components of various frequencies are all delayed in transit along the line by the same delay time $T = \frac{\phi}{2\pi f}$, since $\phi \propto f$ (Compare Chap.3, Sec.15).

In the general case, for a wave travelling on an infinite or properly terminated line, the current i at any point may be given in terms of the current i_s at the sending end, by the equation:-

$$i = i_s \epsilon^{-(\alpha_l + j\beta_l)l}$$

If $i_s = \hat{i}_s \epsilon^{j\omega t}$

then, $i = \hat{i}_s \epsilon^{-\alpha_l l} \epsilon^{j(\omega t - \beta_l l)}$

The voltage is obtained from the relation

$$v = i.z_0$$

The instantaneous values of current and voltage may be written

$$\hat{i}_s \epsilon^{-\alpha \ell} \cos(\omega t - \beta \ell)$$

and

$$\hat{v}_s \epsilon^{-\alpha \ell} \cos(\omega t - \beta \ell + \phi_0),$$

where $\frac{\hat{v}_s}{\hat{i}_s} = Z_0$ (the magnitude of the impedance) and ϕ_0 is the phase angle of the characteristic impedance.

The velocity of propagation is given by:-

$$u = \frac{\omega}{\beta \ell} = \frac{1}{\sqrt{L_\ell C_\ell}}$$

For air-spaced conductors, the value of $c = \frac{1}{\sqrt{L_\ell C_\ell}}$ may be taken as $3 \cdot 10^8$ metres per second. Since $C_\ell \propto K$, the velocity of propagation u along transmission lines using a non-magnetic dielectric of constant K is given by:-

$$u = \frac{3 \cdot 10^8}{\sqrt{K}}$$

The factor $\frac{1}{\sqrt{K}}$ is called the velocity constant and is the ratio $\frac{u}{c}$, where u is the velocity in the feeder concerned and c the velocity if the dielectric were replaced by air (or, strictly, vacuum).

21. Vectorial Representation

Fig. 139 shows the difference between the transmission of sinusoidal waves along lossy and loss-free lines, current vectors only being drawn.

For the loss-free line, \vec{v} and \vec{i} are in phase, and both trace out circles. For the lossy line, there is in general a phase difference ϕ_0 between \vec{v} and \vec{i} , and they trace out similar equiangular spirals.

The curves for the loss-free line should be compared with the vectorial representation of the voltages and currents transmitted by a low-pass filter (Fig. 76). They will be seen to be consistent with the idea that the line is the limiting case of an infinite number of infinitesimal low-pass filter sections. With the transition from the polygon to the circle for all frequencies, there is the elimination of the cut-off frequency and the distortion which it introduces.

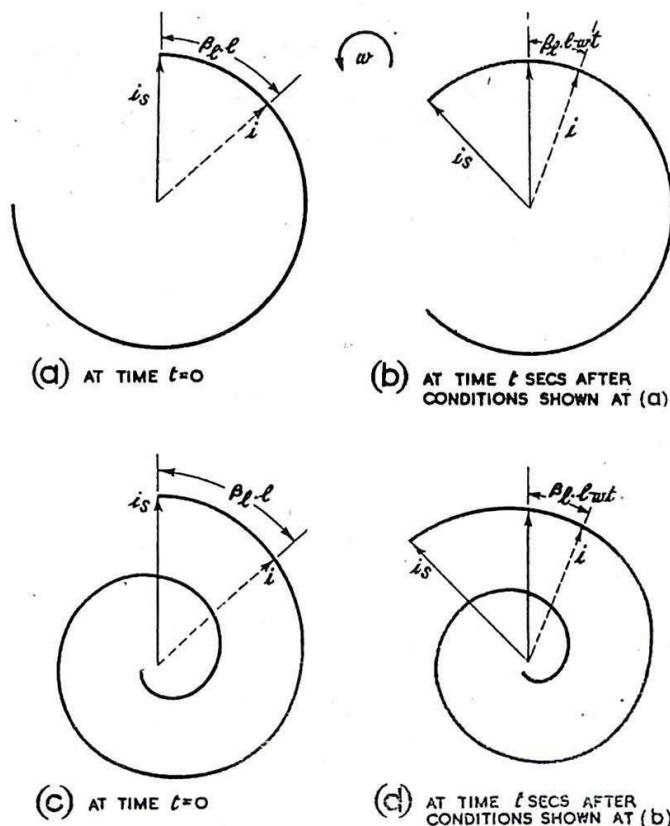


Fig. 139 - Variation of current (or voltage) vectors with line length for loss-free line - (a) and (b) - and lossy line - (c) and (d).

LINES WITH LOW LOSSES

22. Effects on Line Characteristics

If $\frac{\omega L_l}{R_l} \gg 1$ and $\frac{\omega C_l}{G_l} \gg 1$ losses are very small and certain close approximations can be made to the general formulae of Secs. 17 to 21.

The velocity of propagation u is approximately equal to $\frac{1}{\sqrt{L_l C_l}}$ for all frequencies satisfying the above inequalities, the fractional error being approximately $(\frac{R_l}{\omega L_l} + \frac{G_l}{\omega C_l})^2$.

This approximation is justified for all practical purposes in the transmission lines used in radar. It reduces the characteristic impedance z_0 to

$$\sqrt{\frac{L_l}{C_l}} \left\{ 1 + \frac{j}{2} \left(\frac{G_l}{\omega C_l} - \frac{R_l}{\omega L_l} \right) \right\}$$

and it is usually sufficiently accurate to neglect the reactive portion and assume that:-

$$z_0 = \sqrt{\frac{L_l}{C_l}}, \text{ so that it is generally denoted by } R_0.$$

Similarly the loss constant G/l reduces to:-

$$\frac{1}{2} \left\{ R_l \sqrt{\frac{C_l}{L_l}} + G_l \sqrt{\frac{L_l}{C_l}} \right\}$$

$$= \frac{1}{2} \left\{ \frac{R_l}{R_0} + G_l R_0 \right\}.$$

The behaviour of R_l and G_l , and methods of minimising them, are discussed in Sec. 42.

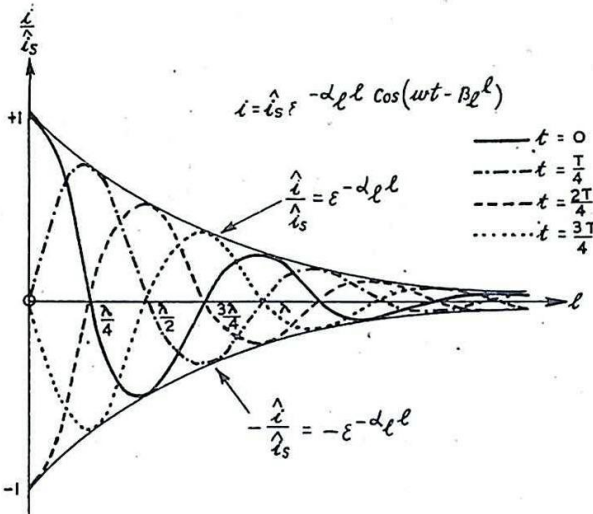


Fig. 140 - Variation of current (or voltage) with distance along the line at selected instants.

23. Effects of Slight Losses on Travelling Waves

Since distortion is usually negligible for the types of line used in radar, the attenuation introduced by low losses is the principal effect to be considered.

The amplitude of a travelling wave decreases as the wave progresses, as illustrated in Figs. 140 and 141(a). This is more simply shown in Fig. 141(b), where the peak values at increasing distances along a lossy line are compared with those for a loss-free line.

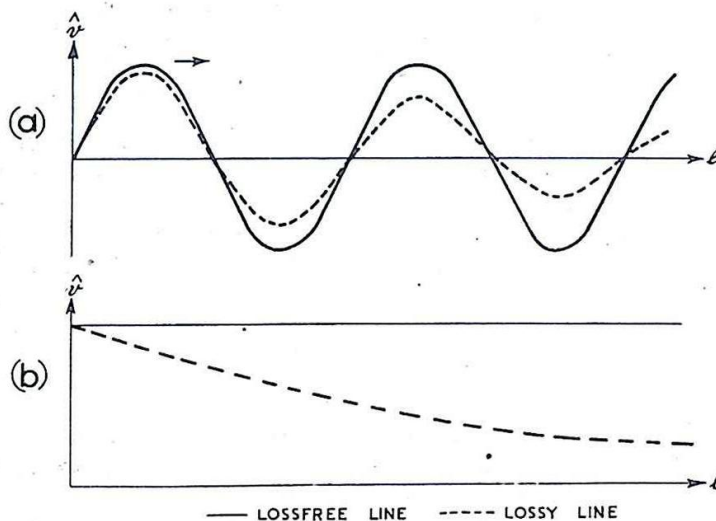


Fig. 141 - Decrease in amplitude of travelling sinusoidal wave due to losses.

24. Effects of Slight Losses on Standing Waves

A simple picture of line conditions can be built up by the method indicated in Figs. 142 and 143. In these figures the attenuation is accentuated to clarify the effects. An open-circuit termination has been chosen. These diagrams will serve also for the short-circuit termination with minor alterations (compare Sec. 15). Other terminations may be considered in the same manner, but as the quantitative results are more complicated, these will be derived analytically where required.

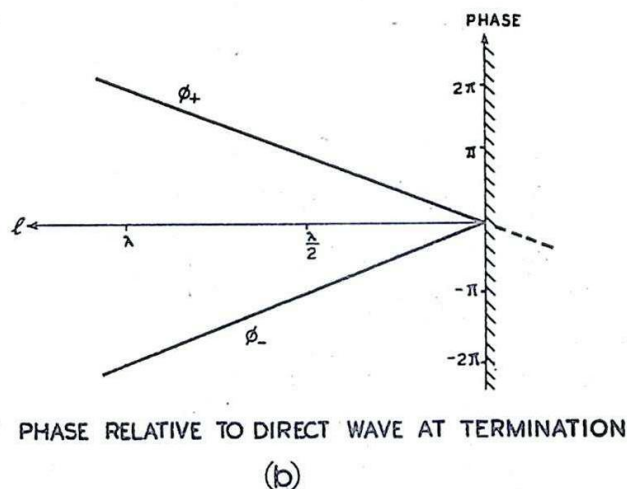
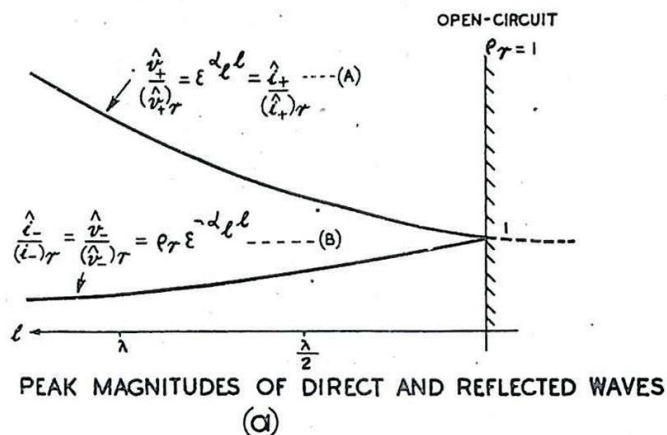


Fig. 142 - Effect of line losses on direct and reflected waves: open-circuited termination.

Because of the reduction in amplitude of the reflected wave compared with the direct wave as the point considered recedes from the termination, the line gives less indication of a mismatch. For very long lines, the amplitude of the reflected wave is negligible compared with that of the direct wave, and the line "looks like" an infinite one, with input impedance R_0 , and standing wave ratio (SWR) unity. In general the SWR at the input is reduced with increase in line length.

Analytically it is convenient to generalise the term Reflection Coefficient to apply to all points on a line, and include the effects of attenuation as well as reflection. If ρ is the reflection coefficient at a distance l from the termination,

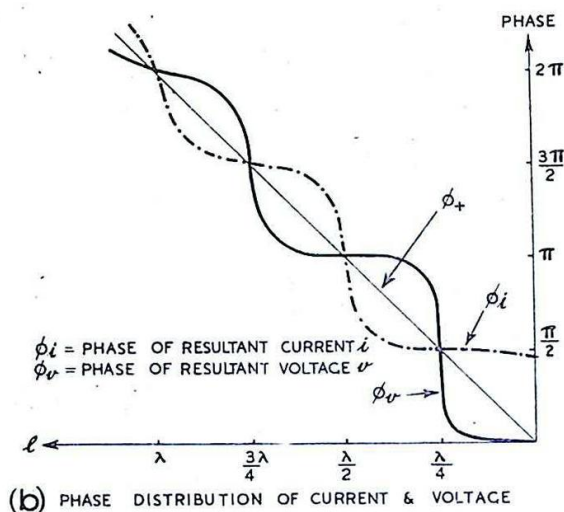
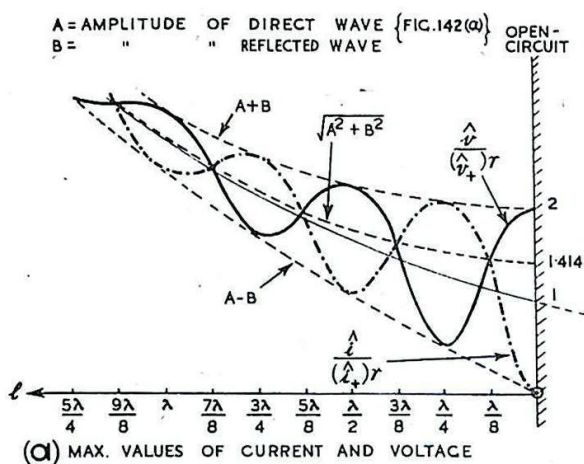


Fig. 143 - Effect of line losses on standing wave: open-circuited termination.

$$\rho = \frac{v_-}{v_+} = \frac{(v_-)_r \epsilon^{-\alpha_l l} \epsilon^{-j\beta_l l}}{(v_+)_r \epsilon^{\alpha_l l} \epsilon^{j\beta_l l}} = \epsilon^{-2\alpha_l l} \rho_r \cdot \epsilon^{-2j\beta_l l},$$

where ρ_r is the value of ρ at the receiving end. (See Sec. 13).

Hence $|\rho| = \epsilon^{-2\alpha_l l} \cdot |\rho_r|.$

At this point the standing wave ratio is given by:-

$$\text{SWR} = \frac{\frac{\hat{v}_+}{\hat{v}_-} + \frac{\hat{v}_-}{\hat{v}_+}}{\frac{\hat{v}_+}{\hat{v}_-} - \frac{\hat{v}_-}{\hat{v}_+}} = \frac{1 + |\rho|}{1 - |\rho|} = \frac{1 + \epsilon^{-2\alpha_l l} |\rho_r|}{1 - \epsilon^{-2\alpha_l l} |\rho_r|}.$$

It follows that as $l \rightarrow \infty$, the SWR $\rightarrow 1$.

Provided α_l is small, it is usually sufficiently accurate to take the SWR as equal to the ratio of voltages at an antinode and a node in the neighbourhood of the point considered.

25. Energy Losses

The energy dissipated in the conductor resistance or dielectric conductance is made manifest in the line by the heat generated. This may have subsidiary effects, such as mechanical distortion of the line, and consequent interference with its electrical properties.

The power loss in a low-loss line of length l may be written:-

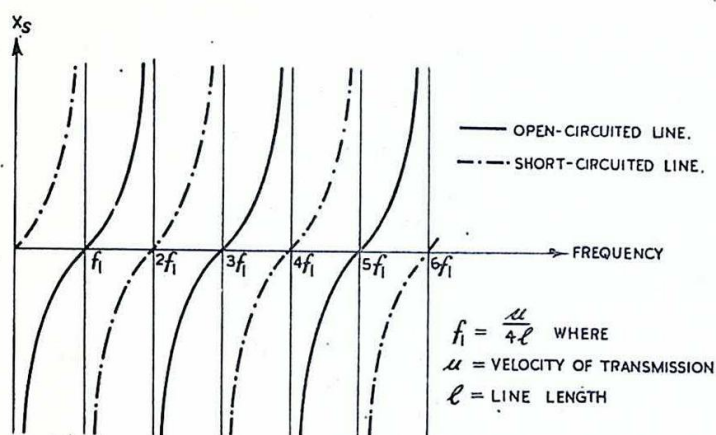
$$20 \log_{10} \epsilon^{\alpha l} + 10 \log_{10} \left[1 + \frac{(S-1)^2}{4S} (1 - \epsilon^{-4\alpha l}) \right] \text{db.}$$

This expression shows the extent to which losses are increased if standing waves are present.

RESONANT LINES

26. General Nature of Resonance in Lines

It follows from the results stated in Secs. 8 - 25 that as the frequency of a signal applied to a short length of transmission line is varied, the line exhibits the properties of series and parallel resonance at specific frequencies. For a uniform loss-less line with an open-circuited termination, the input impedance is either a pure reactance or else is zero or infinite; the variation of this reactance with frequency is illustrated in Fig. 144. The harmonic nature of the resonance characteristics is evident from the figure.



If l is the length of the line, series or parallel resonance occurs, alternately, at all frequencies which are multiples of $\frac{u}{4l}$, where u is the velocity of propagation.

For lines with slight losses, the resonant properties are far more pronounced for shorter lengths; as the line length increases, the input impedance tends more and more to be independent of the line length and to approach R_0 . This is indicated in Fig. 145. Here it is the magnitude of the impedance which is plotted, at a given frequency, for different lengths of line.

A short-circuited line of length l or an open-circuited line of length $2l$ presents a high impedance to a generator of frequency $f = \frac{u}{4l}$, and a reactive impedance to other frequencies.

The impedance is reduced substantially for small deviations of the frequency from resonance. Such a line may be used as a parallel

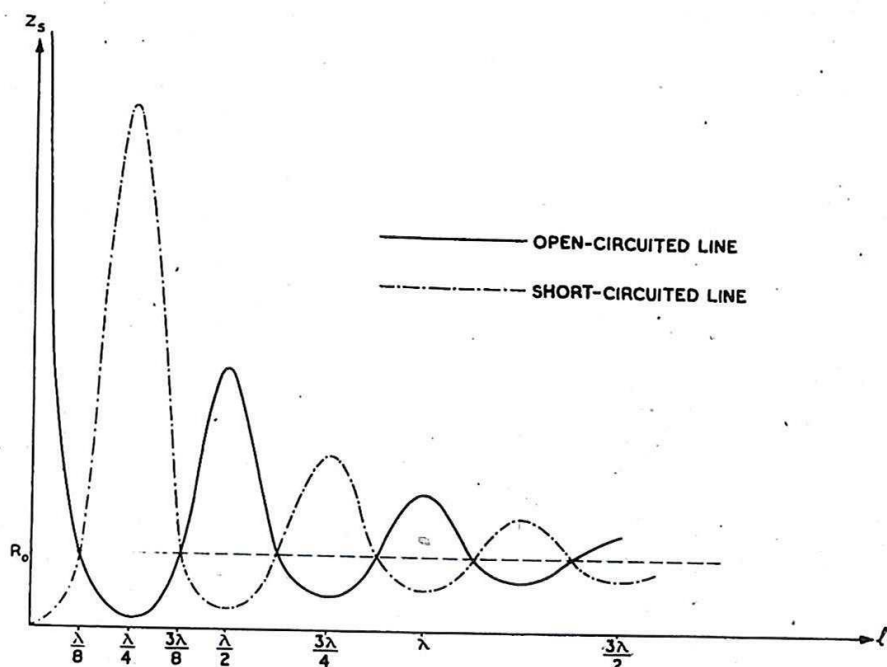


Fig. 145.- Lossy line: variation of magnitude of impedance with line length.

resonant circuit. Similarly, an open-circuited line of length l or a short-circuited line of length $2l$ may be used as a series resonant circuit for a generator of frequency $\frac{u}{4l}$.

For a generator of given frequency f , corresponding to a wavelength $\lambda = \frac{u}{f}$, the resonant lengths are given by the equation

$$l = \frac{m\lambda}{4} \text{ where } m \text{ is an integer.}$$

Non-resonant lines may be used as reactances. The variation of reactance with frequency for open-circuited and short-circuited low-loss lines is illustrated in Fig. 146. In this diagram it is assumed that losses per unit length are independent of frequency.

It should be noted that midway between series and parallel resonant lengths the magnitude of the impedance is always R_0 .

27. Q - Factor of a Resonant Length of Line.

It is known that for a series $L - C - R$ circuit near resonance the magnitude of the impedance is given by $Z = R \sec \phi$ where $\tan \phi = 2Q\delta$, δ being equal to $\frac{f-f_0}{f_0}$, where f is the frequency and f_0 the

resonant frequency. (The error in this approximation is of the order $\frac{1}{Q^2}$.) (See Chap. 1, Sec. 19).

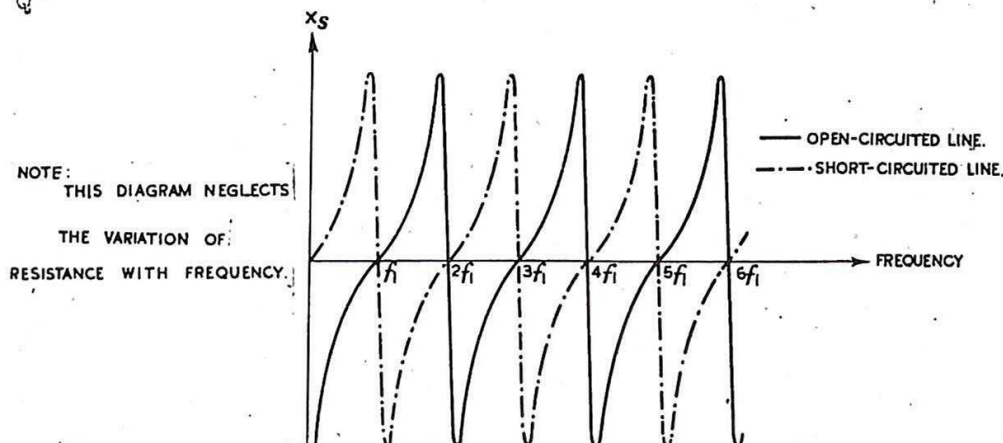


Fig. 146 - Open-circuited lossy line: variation of input impedance with line length.

It may be shown that for an open-circuited low-loss transmission line, approximately $\frac{\lambda_0}{4}$ in length,

$$Z \doteq (R_0 \frac{\alpha_l \lambda_0}{4} \sec \phi),$$

where $\tan \phi = \frac{2\pi\delta}{\alpha_l \lambda_0}$.

It follows that an open-circuited transmission line of length $\frac{\lambda_0}{4}$ behaves like a series tuned circuit having a Q-factor of magnitude

$$Q = \frac{\pi}{\alpha_l \lambda_0}.$$

This reduces to $Q = \frac{\omega L_l}{R_l}$ if dielectric losses are neglected, corresponding to the Q-factor of a coil in lumped-circuit theory.

Similar results are obtainable for short-circuited lines, a quarter-wave-length line corresponding to a parallel resonant circuit with the same value for Q as for the series circuits. Longer lines, $\frac{\lambda_0}{2}, \frac{3\lambda_0}{4}$, etc., have the same Q-factor as the $\frac{\lambda_0}{4}$ line, but not the same dynamic impedance. The longer the length of line (in quarter wave-lengths) the more nearly does the dynamic impedance approach R_0 . For example, a $\frac{\lambda_0}{2}$ short-circuited line has approximately twice the dynamic impedance of a $\frac{\lambda_0}{4}$ open-circuited line.

In practice very large values of Q may be achieved, of the order 10^4 .

28. Lecher bars

Short lengths of line may be used as tuned circuits, often called Lecher Bars. These are usually short-circuited sections of tubular line, the short-circuit termination providing mechanical rigidity with very low loss.

Fig. 147 shows an arrangement which is suitable for use in a tunable oscillator circuit. The curvature of the lines enables tuning to be performed by a rotary motion, varying the effective length of the lechers. For very high frequencies a self-screening type of tuned circuit is required, as unscreened lines become less efficient as the frequency rises (see Sec. 46). This may take the form of a short length of coaxial line. An arrangement suitable for use in a valve circuit is illustrated in Fig. 148. The line is tuned by moving the plunger which may be fitted with a vernier screw for accurate adjustment.

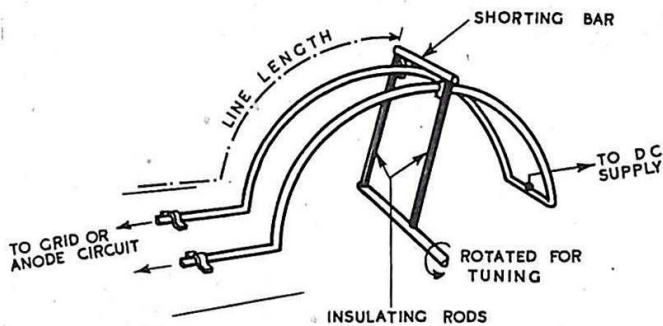
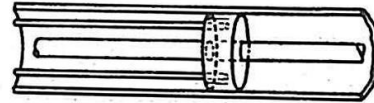


Fig. 147 - Lecher bars for use in oscillators.

Fig. 148 - Adjustable coaxial short-circuited line with plunger tuning.



29. Metallic Insulators

One of the chief difficulties encountered in the use of open-wire feeders is the maintenance of rigid spacing and positioning of the conductors whilst avoiding high dielectric losses. One method, applicable to a single-frequency transmission system, avoids dielectric losses altogether by the use of metallic insulators. The arrangement is illustrated in Fig. 149(a). Such a short-circuited quarter-wave line presents a high impedance to the feeder at the points of junction, and the effect of this on a low impedance line is usually negligible. A frequency variation of $\pm 7.5\%$ is usually permissible before the line becomes appreciably mis-matched.

The problem of supporting the system as a whole does not arise in self-screening lines, such as coaxial systems, where the fields are confined to the inside of the outer conductor and the outside may be earthed at any point without disturbing the electrical properties of the system. But the equivalent problem of supporting the inner conductor may be solved in the same way, by the use of quarter-wave sections of short-circuited line. The arrangement is shown in Fig. 149(b).

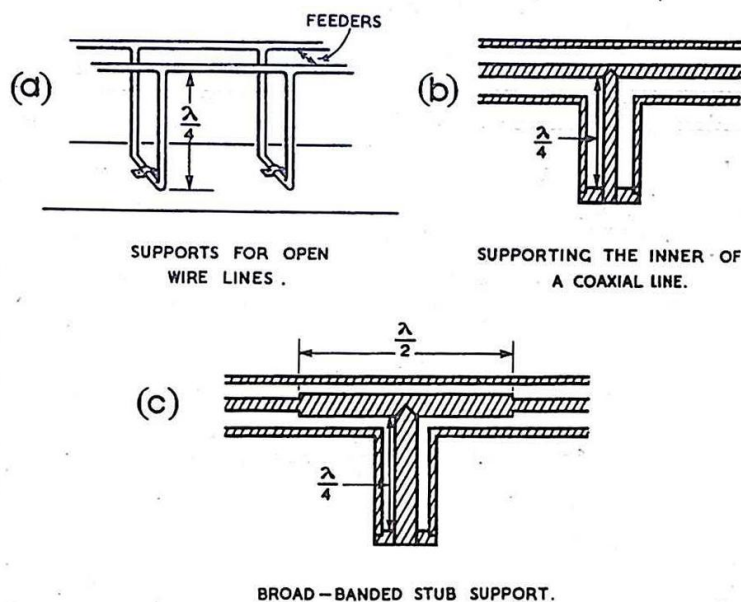


Fig. 149 - Metallic insulators.

A modification to improve the band width of the insulator is shown in Fig. 149(c). The thickening of the inner conductor forms in effect two $\frac{\lambda}{4}$ transformers back to back, in parallel with the $\frac{\lambda}{4}$ insulator at the junction. The analysis of this arrangement is dealt with in Sec. 54 (viii).

30. Quarter-Wave Sleeve Rejector (Rotating Joint)

This is a method of using a $\frac{\lambda}{4}$ open-circuited coaxial line to join electrically two other sections. Ideally the impedance presented by the device is zero at the frequency of operation.

The arrangement is shown schematically in Fig. 150(a). This represents a $\frac{\lambda}{4}$ section inserted in series with an open-wire feeder.

Since the $\frac{\lambda}{4}$ line has zero input impedance it does not affect the flow of energy along the transmission line.

If the figure is rotated about the lower conductor, the upper conductor generates the outer of a coaxial cable, and the $\frac{\lambda}{4}$ line takes the form of a double circular plate of radius $\frac{\lambda}{4}$ with the same properties as in the balanced pair arrangement; this stage is shown in cross-section in Fig. 150(b). If now the protruding $\frac{\lambda}{4}$ "line"

is folded back on one of the outers the line presents the appearance of Fig. 150(c). The separate portion AB then becomes redundant, since the outer surface of the cable CB is able to fulfil its functions; when it is removed, the circuit appears as in Fig. 150(d). Finally, the inner may be subjected to the same process as the outer, the complete joint being depicted in Fig. 150(e). Such an arrangement is suitable for a rotating coupling, since there is no mechanical contact to introduce frictional losses.

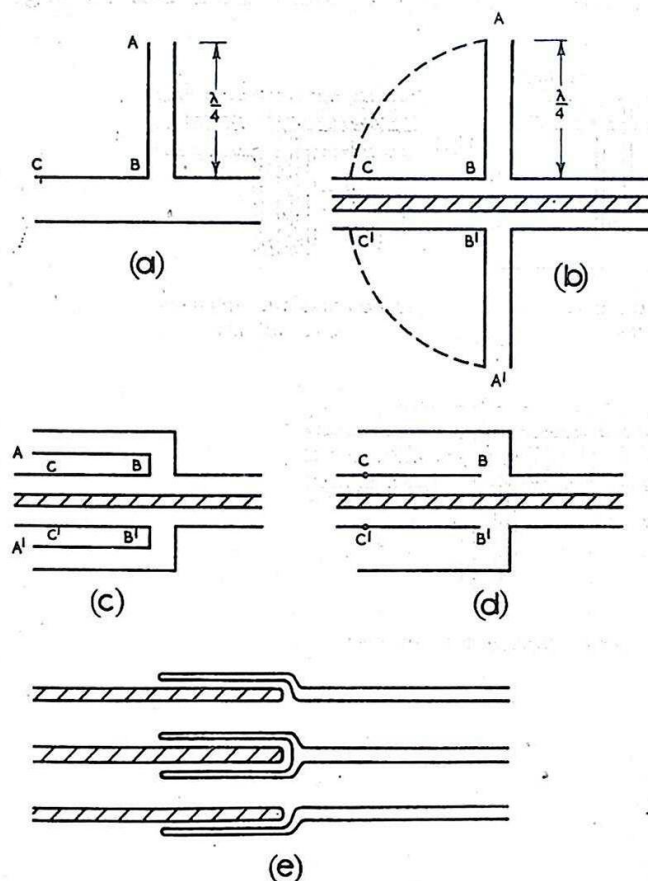


Fig. 150 - Evolution of $\lambda/4$ sleeve rejector (rotating joint).

31. Stub Reactances

Short (usually $< \frac{\lambda}{2}$) lengths of open-circuited or short-circuited transmission line, used as reactances to modify the standing wave distribution on a transmission system, are called Stubs. Open-circuited stubs are frequently used with open-wire feeders where the conductors are rigid metal tubes or bars. For less rigid structures and for coaxial lines, short-circuited stubs are invariably used.

The variation of input reactance with stub length was given in Figs. 131 and 134. Since the stub is usually connected in parallel with the transmission line it is more usual to analyse stub problems in terms of admittances.

The input admittance of a loss-less short-circuited stub of length l is given by

$$y = -j \cot \frac{2\pi l}{\lambda}, \quad (\text{a short length is inductive}).$$

The input admittance of a loss-less open-circuited stub of length l is given by

$$y = j \tan \frac{2\pi l}{\lambda}, \quad (\text{a short length is capacitive}).$$

MATCHING

32. Reasons for Matching

Theoretically it is possible to design a transmission system so that the input and output impedances of all its elements are resistive and of the same value, and matching problems do not arise. However, it is seldom practicable to do this, and more often than not energy is transferred from a generator to an aerial system at several different impedance levels and correspondingly different voltage levels. It has been shown in Secs. 8 to 16 that if a length of transmission line is not terminated in its characteristic impedance standing waves occur on the line, and the input characteristics vary with line length or, alternatively, with frequency. It is inadvisable to have a high standing wave ratio on a long feeder system, since a small change in frequency might cause a large change in input impedance. For example, if a line is 10λ long, a 2.5% variation in frequency would make this 10.25λ or 9.75λ , and this could replace a voltage node at the sending end of the line by a voltage antinode; this would change the input impedance from R_r to $\frac{R_o^2}{R_r}$, or vice versa, which is a variation in

impedance equal to the square of the SWR. A similar change of frequency would have relatively little effect if the standing waves were confined to a short matching section less than one wavelength long.

As pointed out in Sec. 25 losses are bound to be heavier if standing waves are present on a feeder system, owing to the extra losses from the oscillatory energy of the standing waves. In addition, losses usually tend to increase with abnormally high voltages and currents, and at antinodes it is possible for dielectric breakdown or corona discharge to occur, with prohibitive loss of energy.

Standing waves on the main portion of a transmission system are avoided by the use of matching sections which fulfil the purpose described in Chap. 3, Sec. 4. They are inserted between source and load and ensure that the transmission lines are terminated in their characteristic impedances. Matching sections may also be used between two lengths of line which have different characteristic

impedances, or to match a feeder system to a generator.

33. Half-Wave Transformer

This is the simplest matching section. It is equivalent to a 1:1 transformer which does not change the input impedance, but transfers the input terminals to a more convenient position. The choice of length, a multiple of $\frac{\lambda}{2}$, enables a line of any convenient characteristic impedance to be used.

If a low-loss line is used, the formula of Sec. 16 can be employed, viz:

$$\frac{z_s}{R_0} = \frac{z_r + j R_0 \tan \frac{2\pi l}{\lambda}}{R_0 + j z_r \tan \frac{2\pi l}{\lambda}}$$

Since l is a multiple of $\frac{\lambda}{2}$, $\tan \frac{2\pi l}{\lambda} = 0$, so that $z_s = z_r$, irrespective of the value of R_0 .

This principle may be employed to join two similar sections of transmission line by a third section of different characteristic impedance. It is the fundamental principle often employed in the construction of plugs and sockets. These usually introduce sections of line of different characteristic impedance, and if their combined length is made a multiple of $\frac{\lambda}{2}$, standing waves on neighbouring sections are avoided.

Typical joints are shown in Fig. 151.

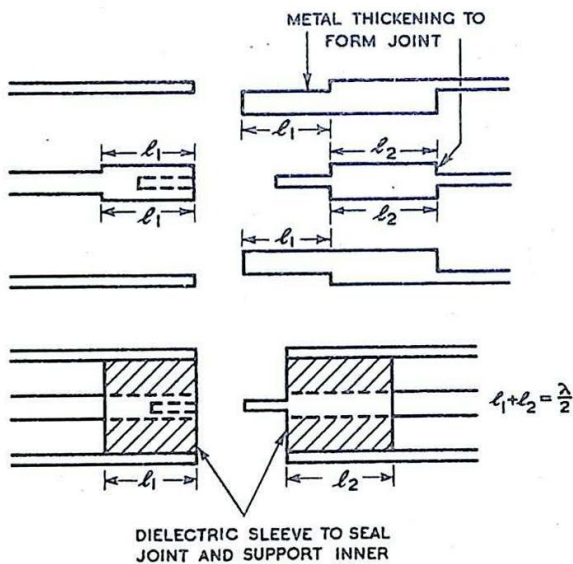


Fig. 151 - Joints designed on the $\frac{\lambda}{2}$ 1:1 transformer principle.

34. Quarter-Wave Matching Sections

These sections of transmission line, odd multiples of $\frac{\lambda}{4}$ in length, are impedance transformers. The formula quoted in Sec. 33 may be written:-

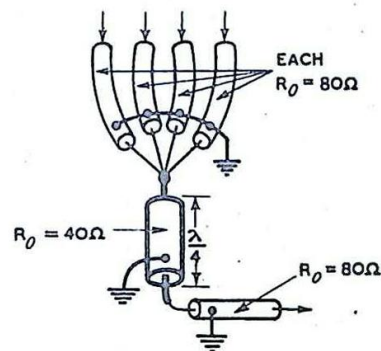


Fig. 152 - Typical matching problem: use of quarter-wave transformer.

$$\frac{z_s}{R_0} = \frac{z_r \cot \frac{2\pi l}{\lambda} + j R_0}{R_0 \cot \frac{2\pi l}{\lambda} + j z_r}$$

Making the substitution $l = \frac{(2r+1)\lambda}{4}$, we have $\cot \frac{2\pi l}{\lambda} = 0$ and $z_s = \frac{R_0^2}{z_r}$. In the simplest case, where the terminating impedance is resistive (R_r), $z_s = \frac{R_0^2}{R_r}$ and is a pure resistance.

The quarter-wave section thus acts as an impedance transformer. Given two lines of characteristic impedance R_1 and R_2 , a $\frac{\lambda}{4}$ section designed to match the one to the other would need to have a characteristic impedance $R_0 = \sqrt{R_1 R_2}$.

A typical case arises in which a standard cable is to be matched to several similar cables which are in parallel at the junction. This is illustrated in Fig. 152. The four 80 ohm cables present an impedance of 20 ohms; to match this to the single cable requires a $\frac{\lambda}{4}$ section of characteristic impedance $\sqrt{20 \cdot 80} = 40$ ohms.

To preserve the initial spacing of the conductors the matching section either must be made of thicker material, or must use a different dielectric with a larger constant, K .

It may be noted that the use of metallic insulators is the limiting case of a $\frac{\lambda}{4}$ matching section where $z_s = \infty$ and $z_r = 0$.

R_0 may have any value.

35. Double Quarter-Wave Line

The frequency sensitivity of a quarter-wave matching section makes it usable at one frequency only. By extending the principle to the use of two quarter-wave transformers in cascade, with appropriate characteristic impedances, a broad-band match may be obtained.

The characteristic impedances to match R_1 to R_2 should be in the order

$$R_1, \frac{1}{2}R_0, 2R_0, R_2 \quad \text{where:}$$

$$\frac{2R_0}{\frac{1}{2}R_0} = \left(\frac{\frac{1}{2}R_0}{R_1}\right)^2 = \left(\frac{R_2}{2R_0}\right)^2$$

This relation is more simply expressed in terms of the logarithms of the ratios of consecutive impedances. These ratios for the three junctions are:

$$\frac{\frac{1}{2}R_0}{R_1}, \frac{2R_0}{\frac{1}{2}R_0}, \text{ and } \frac{R_2}{2R_0} \quad \text{and the logarithms are in the}$$

ratio 1:2:1.

If $R_0 = aR_1$, then ${}_2R_0 = a^3R_1$ and $R_2 = a^4R_1$.

Thus, if R_1 and R_2 are given, a may be determined and hence ${}_1R_0$ and ${}_2R_0$.

To a first approximation this arrangement ensures that the reactive term introduced, by a change of frequency, in the output impedance of the first quarter-wave section, is cancelled by an equal and opposite reactive term in the input impedance of the second quarter-wave section.

The output impedance of the first quarter-wave section can be written:-

$$\frac{z_1}{{}_1R_0} = \frac{R_1 \cot \frac{2\pi\ell}{\lambda} + j{}_1R_0}{{}_1R_0 \cot \frac{2\pi\ell}{\lambda} + jR_1}, \quad \text{where } \cot \frac{2\pi\ell}{\lambda} \text{ is small} = x, \text{ say.}$$

(Cot $\frac{2\pi\ell}{\lambda}$ is zero if ℓ is exactly $\frac{\lambda}{4}$.)

$$\begin{aligned} \text{We then have } \frac{z_1}{aR_1} &= \frac{j a R_1 + xR_1}{jR_1 + a xR_1} \\ &= \frac{a - jx}{1 - ajx} \doteq (a - jx)(1 + ajx) \doteq a + jx(a^2 - 1). \end{aligned}$$

$$\text{So that } z_1 \doteq R_1 a^2 \left\{ 1 + jx (a^2 - 1) \right\}.$$

Similarly z_2 , the input impedance of the second quarter-wave line can be written

$$\frac{z_2}{{}_2R_0} = \frac{R_2 \cot \frac{2\pi\ell}{\lambda} + j{}_2R_0}{{}_2R_0 \cot \frac{2\pi\ell}{\lambda} + jR_2},$$

$$\text{and this reduces to } z_2 \doteq R_1 a^2 \left\{ 1 - jx (a^2 - 1) \right\}$$

This satisfies the matching conditions described in Chap. 3, Sec. 4. An alternative treatment is given in Sec. 54(iv).

The principle of the double quarter-wave line may be extended to any even number of quarter-wave sections, the characteristic impedances being chosen so that the logarithmic ratios at successive junctions form a "binomial coefficient" series, (e.g. 1, 2, 1; 1, 4, 6, 4, 1; etc.) In general an odd number of sections is sensitive to frequency changes; an even number is not. When a large number of sections is used the change in impedance from one end to the other is approximately exponential, and very broad band coverage is afforded.

36. Matching by Stubs

The principle of stub matching is to shunt a section of transmission line by suitable reactances at various points so that the input impedance is made equal to the required value. [Series stubs are not commonly used in transmission line systems.]

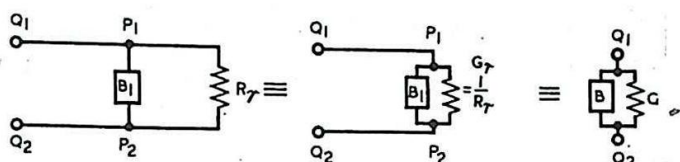


Fig. 153 - Stub reactance in parallel with a matched line.

If a section of correctly terminated line is shunted at some point by a single reactance such as would be presented by a short length of open circuited line, both the input resistance and the input are changed. Fig. 153 illustrates this point. Since we are concerned with parallel circuits it is more convenient to deal in admittances. The section of properly terminated line to the right of $P_1 P_2$ may be replaced by G_r , its characteristic admittance, and the stublength by a susceptance B_1 . As B_1 is varied, the input conductance G and susceptance B are subject to the type of variation indicated in Fig. 154.

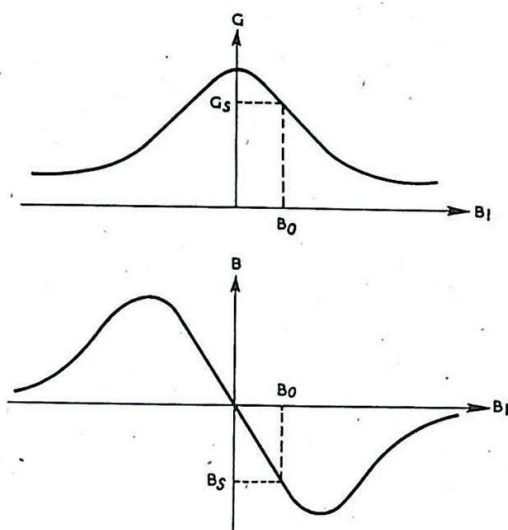


Fig. 154 - Typical variation of B and G with B_1 (see fig. 153).

There is usually a limited range of values over which G varies. If it is required to match the line at $Q_1 Q_2$ to some characteristic admittance G_s such that $\hat{G} > G_s > G$, it is possible to find at least one value of B_1 , (B_0 , say) for which $G = G_s$; corresponding to this value of B_1 there is an unwanted susceptance B_s . By shunting $Q_1 Q_2$ with an equal and opposite susceptance $-B_s$, the input admittance is made equal to G_s . The schematic appearance of this arrangement is shown in Fig. 155, and the corresponding mechanical design in Fig. 156. This method of matching is known as double stub matching.

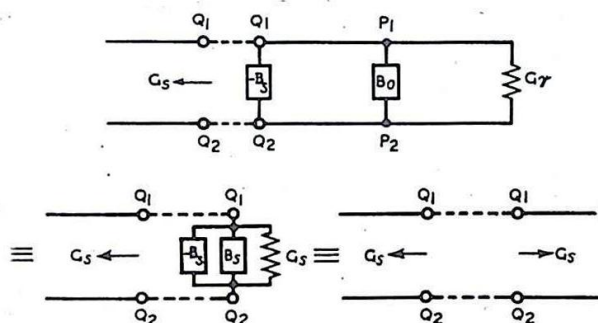
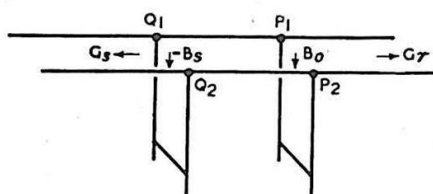


Fig. 155 - Double stub matching: schematic diagram.

Fig. 156 - Arrangement of stubs providing the above matching requirements.



Because of the limited range of values for G it is not possible to obtain a correct match for all values of G_S . There are various methods of overcoming this difficulty. One is to use a third stub. If a match is not possible for one setting of this stub, the setting is changed (usually by approximately $\frac{\lambda}{4}$) and it

it then possible to effect a match using the two other stubs as previously described. Another method is to change the position of the stubs relative to the two lines. This is indicated in Fig. 157. The portion P Q R S is made unsymmetrical and reversible so that if a match is not possible with the arrangement as shown in (a), it is possible in the arrangement (b).

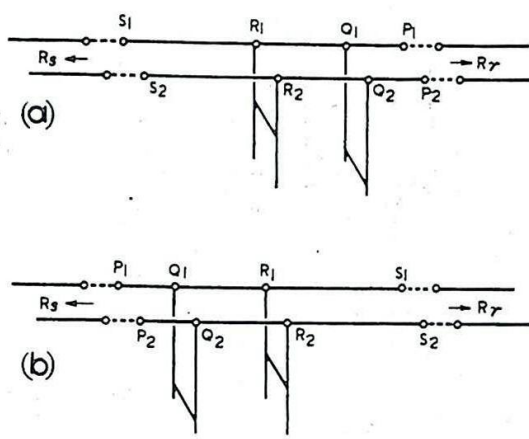


Fig. 157 - Modification to double stub method to extend the matching range.

The single stub method of Fig. 158 is another way of obtaining a match for any relation between source and load impedances. The two variable distances are the stub length l_2 and its distance l_1 from the termination T_1 T_2 . This arrangement is sometimes used with open wire feeders, but it is not readily adaptable to coaxial lines. The disadvantage lies in the mechanical arrangements for the sliding contact. The necessity for robustness makes it difficult to design a sliding contact which does not interfere with the characteristic impedance of the line due to the thickening of the inner conductor caused by the sliding sleeve, as indicated in Fig. 159(a). This does not matter at the short-circuited end of a stub since, on the side remote from the line, there are no restrictions on robustness, as illustrated in (b). Further trouble arises through the necessity for a slot along which the inner conductor of the coaxial stub can move. This is discussed more fully in Sec. 39. Alternatively the same effect as the sliding stub may be created by inserting a line lengthener between the stub and termination; but such an arrangement is not commonly used because it is seldom mechanically convenient and is not without its own sliding joint troubles.

Stub matching is further considered in Sec. 54(v) and (vi).

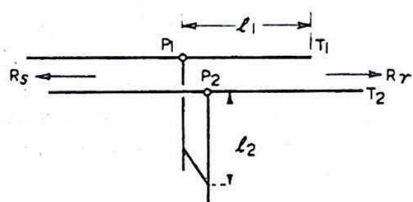
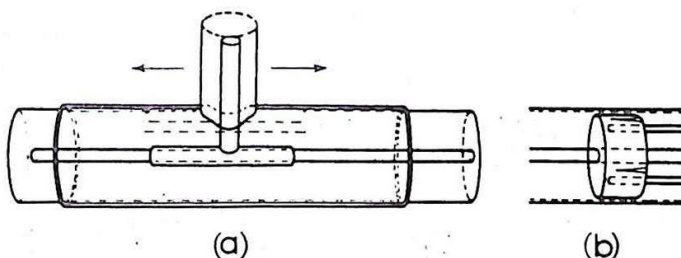


Fig. 158 - Single stub matching arrangement.

Fig. 159 - Sliding joints: movable coaxial stub and plunger.



37. Slugs

Slug is the term used to describe a device which deliberately produces a local variation in the characteristic impedance of a transmission line. In practice, it may take the form of a thickening of one of the conductors of a coaxial cable, or of both conductors of a balanced pair, usually by a movable sleeve. The method is illustrated in Fig. 160. The sleeves may be made either of conducting material or of some dielectric material different from the main dielectric.

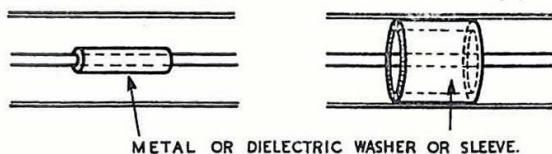


Fig. 160 - Slugs: typical arrangements.

Slugs are usually, but not necessarily, an odd number of quarter-wave-lengths long. In this case they act similarly to the quarter-wave matching sections already described. Movement of the slug along the line provides one degree of freedom for varying the input impedance. As pointed out in Sec. 36, two degrees of freedom in the matching device are required, so that a single slug is not by itself an adequate matching device. Double slug matching is dealt with in Sec. 54(vii).

38. Balance to Unbalance Transformer

This term is used to denote a device for matching an unbalanced line to a balanced load or source (see Chap. 3, Sec. 1). Fig. 161 shows an unbalanced line connected to a balanced load R_L . It is clear that if the lower conductor were to be earthed at all points, the lower half of the load would be short-circuited. What is required is a four terminal network, arranged as in Fig. 162, which enables the lower conductor to be earthed but provides the load with a balanced feed.

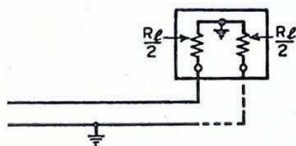


Fig. 161 - Matching unbalanced line to balanced load.

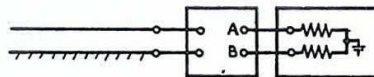


Fig. 162 - Balance to unbalance transformer.

One method uses a series Half-Wave Loop, shown in Fig. 163. The voltage and current at B are equal (neglecting losses) and anti-phase to those at A, which is the condition required for feeding the load in push-pull.

The impedance presented to the line at A E_1 by the rest of the circuit is $\frac{R_L}{2}$, between A and E_2 , in parallel with the input impedance of the half-wave loop. This also is $\frac{R_L}{2}$, since the loop, terminated in $\frac{R_L}{2}$, acts as a 1:1 transformer. Hence the resultant impedance is $\frac{R_L}{4}$ and, to avoid reflection, this should be the characteristic impedance of the main feeder. For single-frequency systems the characteristic impedance of the half-wave loop is immaterial, as shown in Sec. 33.

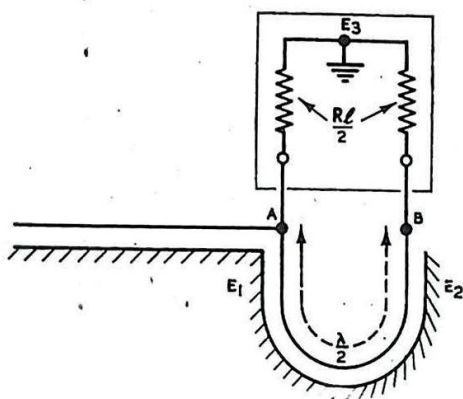


Fig. 163 - Use of half-wave loop.

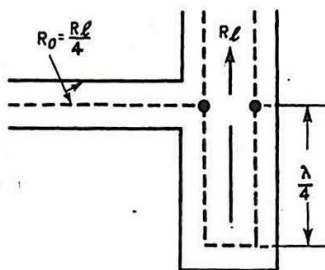


Fig. 164 - Matching coaxial cable to screened pair.

The arrangement of this loop method for coupling a coaxial cable to a screened balanced pair is shown in Fig. 164.

The transformer may be tuned to a required frequency by fitting a variable extension to the loop, as in a trombone. Such an arrangement is sometimes called a "Trombone Matching Section".

Another common method of satisfying the requirements is to use a quarter-wave "can", "skirt" or "balun". The schematic arrangement for an open-wire circuit is shown in Fig. 165. The lower conductor at (a) is earthed at the source, but cannot be at earth potential at all points without short-circuiting the lower half of the load. There is no reason, however, why there should not be a section GC, of the lower conductor, which is at earth potential, while standing waves are present on the remainder, BC, as shown at (b).

Standing wave currents i_3 and i_4 are present on the line as indicated. C is a voltage node of this standing wave system, which exists between BC and earth. There is a similar standing wave between the corresponding portion of the upper conductor and earth, and a current i_1' will flow in the upper conductor due to this standing wave, and an equal and opposite current i_4' will flow between E and D.

Denoting the travelling wave currents by i_1 and i_2 , as indicated in Fig. 165(b), it follows from elementary circuit theory that

$$i_1 = i_2 \quad \text{and} \quad i_3 = i_4.$$

The value of the current at a particular point is indicated by the use of a prefix; e.g. the currents in the two halves of the load are i_1 , i_2 respectively whilst the value of i_3 at B is denoted by i_3 .

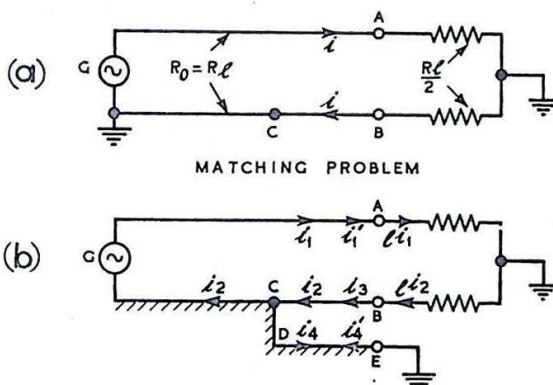


Fig. 165 - Current in standing wave section.

For equal currents in the two halves of the load, i.e.

$$I_1 = I_2,$$

$B I_3$ must be equal to $A I_1$.

The coaxial arrangement may be considered as generated by rotating the diagram of Fig. 165 (b) about the upper conductor, to give the arrangement shown in Fig. 166. Provided the main outer cable is a perfect screen, this ensures that there is a very high impedance between the inner conductor and earth, since there is no coupling between the currents in the inner conductor, and outside the outer. The currents corresponding to i_1 and i_4 in Fig. 165(b) are therefore negligible.

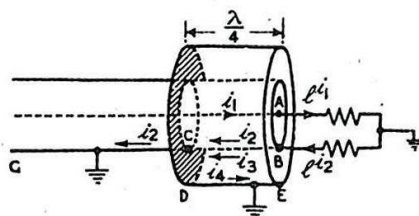


Fig. 166 - Matching can.

It follows that the result deduced above, namely

$$B I_3 = A I_1$$

reduces, in the coaxial line, to

$$B I_3 = 0.$$

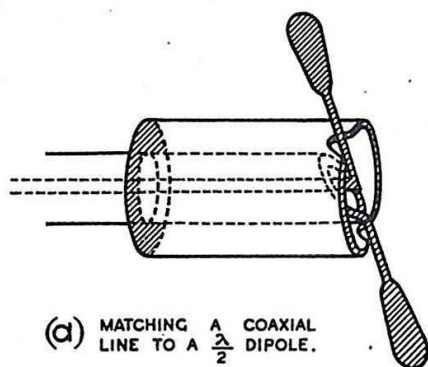
The standing wave between B and C thus has a voltage node at C and a current node at B. Hence BC is an odd number of quarter-wave lengths; usually this is made $\lambda/4$, the outer portion of the can being earthed so that points C and D are at earth potential.

In the coaxial arrangement the currents i_2 and i_3 are separated, flowing in the inner and outer surfaces respectively of the outer conductor BC; i_4 flows in the inner surface of the can DE. The outer surface of the can and the remainder of the cable CG may be earthed everywhere.

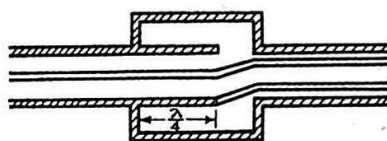
In this type of balance-to-unbalance transformer the characteristic impedance of the line should be made equal to the resistance of the load.

The use of a $\lambda/4$ can is especially suited to feeding a half-wave dipole from a coaxial line, as shown in Fig. 167(a). It is preferable to the half-wave loop system, particularly if the aerial is to be "spun", and mechanical symmetry is desirable. An extension of the arrangement for joining a screened balanced pair to a coaxial cable is shown in Fig. 167(b).

The chief disadvantage of both the transformer systems



(a) MATCHING A COAXIAL LINE TO A $\frac{\lambda}{2}$ DIPOLE.



(b) MATCHING COAXIAL CABLE TO SCREENED PAIR.

Fig. 167 - Matching a coaxial line to (a) $\lambda/2$ dipole and (b) screened pair.

so far described is their sensitivity to frequency changes. A change of frequency causes a mismatch and unbalance in both cases. In the half-wave loop method the input impedance of the loop, terminated in $\frac{R_l}{2}$ is not $\frac{R_l}{2}$ except at the frequency of operation

unless that is its characteristic impedance. In any case, the antiphase relation no longer holds, and a phase mismatch is unavoidable; i.e., the currents in the twin conductors are not antiphase. This changes the input impedance and destroys the correct termination of the coaxial line. In the can method, if the line BC of Fig. 166 is not $\frac{\lambda}{4}$, B is not a current node, so that i_{B3} is not zero.

This makes i_{i2} different from i_{i1} and unbalances the load. Also, the input impedance z_{BE} is not infinite but is reactive and appears in parallel with the lower half of the load resistance, resulting in a mismatch.

It is possible to extend the second method to avoid appreciable unbalance and phase variations at the junction over a wide frequency band.

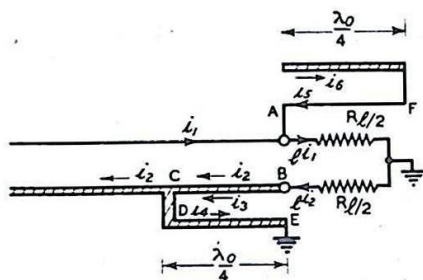


Fig. 168 - Use of $\lambda/4$ stub to remove unbalance in load.

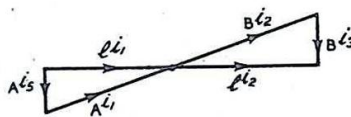


Fig. 169 - Reactive currents (see fig. 168).

A method which permits phase variations but provides a reasonably balanced wide-band transformation is illustrated in Fig. 168. In parallel with the upper half of the load is placed a reactance due to the stub length AF equal to that of DE, in parallel with the lower half, so that the current balance is preserved. This is illustrated in the vector diagram of Fig. 169, the currents referring to those indicated in Fig. 168. The vector relationship indicated implies that the line AF is the same length, and has the same characteristic impedance as the original stub DE.

The actual arrangement is shown in Fig. 170 where a coaxial line is matched to a screened parallel pair. The two stubs AF and BD have the same dimensions, the length $\frac{\lambda_0}{4}$ being the quarter wavelength for the middle of the frequency band over which matching is required. The distance ΔL must be negligible compared with λ_0 .

The Pawsey stub, illustrated in Fig. 171, is based on the same principle as this wide-band matching device, but dispenses with the screening can. The currents flow as indicated in the figure, and correspond to those shown in Fig. 170. The place of the inner surface of the can is taken by neighbouring earthed conductors, carrying the standing wave currents i_4 and i_6 , and if these are remote, the effective impedance will be high, so that the currents i_3 , i_4 , i_5 and i_6 are small. These currents will not, however, be zero, and some radiation is inevitable from the standing wave developed between the stub and the coaxial outer. In practice the short-circuiting plate is adjusted until the best possible match is obtained.

interfere with the mid-frequency current distribution of the solid stub method, but changes the matching arrangements as shown in Fig. 176. If R_2 is the characteristic impedance of the compensating stub, the condition for optimum compensation is $R_0^2 = 2R_1 R_2$. Provided $R_1 \gg R_0$, this circuit is an excellent transformer.

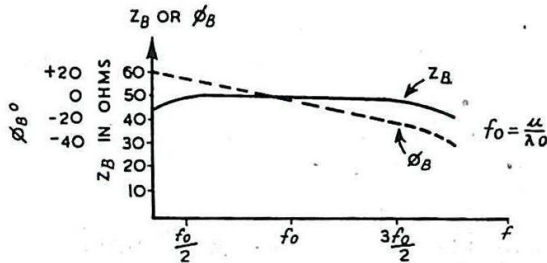


Fig. 173 - Variation of impedance and phase with frequency for wide-band stub match.

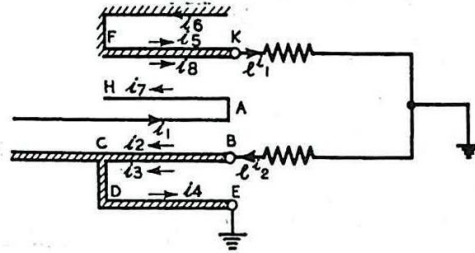


Fig. 174 - Use of auxiliary open-circuited stub as series compensating reactance.

The calculated impedance of a transformer where $R_f = R_0 = 50\Omega$, $R_1 = 100\Omega$ and $R_2 = 12.5\Omega$, is shown in Fig. 177, where the capacitance across Δl is neglected.

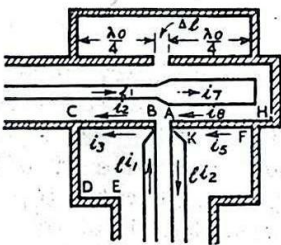


Fig. 175 - Construction of wide-band matching device using compensating open-circuited stub.

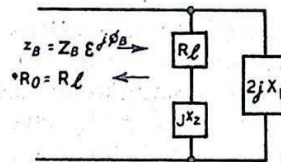


Fig. 176 - Equivalent circuit for Fig. 175.

This type of transformer is specially suited for a transfer from a stationary to a rotating member. Since the coaxial line does not make contact with any part of the rest of the system, it can be kept stationary while the whole transformer and balanced two-wire line rotates around it.

39. Standing Wave Indication

If a feeder system is not properly terminated standing waves will occur with the disadvantage described in Sec. 32. To avoid these it is usually essential at UHF

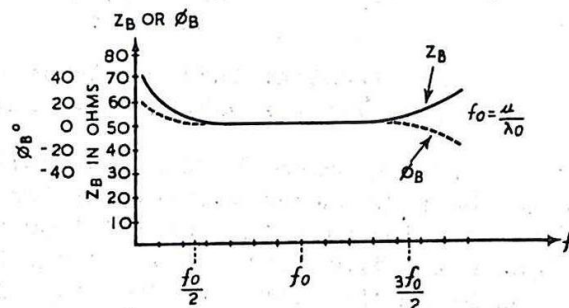


Fig. 177 - Variation of impedance and phase with frequency for wide-band stub match with compensating open-circuited stub.

to use empirical means, since it is not ordinarily possible to coordinate manufacture and design to a degree of accuracy sufficient to ensure the required matching conditions.

Standing wave indicators, which show the variation of field strength with line length, are readily usable with open-wire lines. Coaxial lines introduce considerable difficulties, since it is impossible to measure the field strength inside the cable without making a slot in the outer conductor for the insertion of a probe (or loop). The actual insertion of this probe and its movement along the slot change the line characteristics and increase the difficulty of detecting true standing waves, i.e., those which do not depend for their existence on the presence of the probe or the slot. The same problem exists to a lesser degree in open-wire systems, but usually the presence of a small lightly loaded pickup probe or loop has little effect on the line characteristics. If miniature technique could be sufficiently developed, some improvement might be effected with coaxial or screened lines, but the difficulty is that the smaller the conductors become, the weaker must be the field strength, otherwise there would be a breakdown of the dielectric, or corona discharge. Since at frequencies of the order of 3000 Mc/s and upwards amplification is not yet practicable, the low power available from small pick-ups is the limiting factor in standing wave indication. It is therefore necessary at these frequencies to exercise the greatest care in design to minimise the likelihood of mismatching. The only empirical check practicable is usually an overall one, in which stubs or transformers are adjusted to provide maximum power delivered to the load, usually the aerial system.

The essential feature of a standing wave indicator is a galvanometer measuring field strength. In one of the most elementary forms this is a simple neon lamp, the brightness of which increases with the intensity of the alternating electric field in which it is placed. A more complicated indicator might consist of a pick-up probe or loop coupled to a resonant circuit across which is placed a resonance indicator, such as a valve voltmeter. This is more accurate than the neon indicator, and, since it absorbs but little energy, can be designed to have negligible effect on the line under test. The indicator is moved along the line with a constant disposition relative to the conductors, and the meter or neon shows the increase or decrease in intensity. Matching devices are then adjusted until the standing wave ratio is a minimum.

If the meter is a square-law device, as is often the case, it is the square of the standing wave ratio, or as it is sometimes called, the Power Standing Wave Ratio, which is obtained directly from maximum and minimum readings.

40. Common T/R Circuits

Radar systems which use a common aerial for transmitting and receiving require a special type of transmission system which ensures that the signal energy takes the correct path on each occasion. Two basic circuits, the series and parallel combinations, are depicted in Fig. 178. In both arrangements it is desirable that, to avoid standing waves, the characteristic impedances of the three cables at the junction should have the same value, R_0 .

In the parallel circuit, ideally:-

when transmitting, $z_A = R_0, \quad z_T = R_0, \quad z_R = \infty;$

when receiving, $z_A = R_0, \quad z_T = \infty, \quad z_R = R_0.$

In the series circuit, ideally:-

when transmitting, $z_A = R_0$, $z_T = R_0$, $z_R = 0$;

when receiving, $z_A = R_0$, $z_T = 0$, $z_R = R_0$.

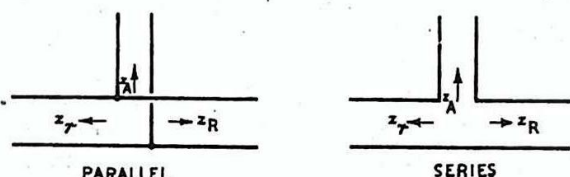


Fig. 178 - Alternative arrangements for common T/R working.

In addition, in the series case the spacing of the lines must not be such as substantially to increase the path lengths of one conductor compared with the other either when transmitting or when receiving. [This amounts to the normal requirement that the spacing of the conductors must be small compared with the wavelength.]

These changes in impedance must be synchronised with the firing of the transmitter. The change in z_T may be due merely to the transmitter ceasing to operate. If the impedance which it presents at the junction changes from R_0 to either a sufficiently high or a sufficiently low value when oscillations cease, one of the requirements of the basic circuits is thereby automatically satisfied. The other requirement necessitates special provision to change the input impedance to the receiver line.

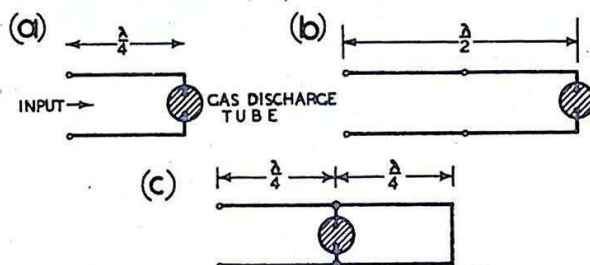


Fig. 179 - Basic switching circuit and modifications.

A basic circuit for providing this change is shown in Fig. 179. A $\frac{\lambda}{4}$ section of transmission line is terminated in a spark gap

or other discharge valve. The voltage of the received pulse is quite inadequate for igniting this valve, so that its impedance remains very large during reception. When the transmitter fires, the pulse ignites the valve which then possesses a small impedance, which is maintained by the ionized gases without absorbing further appreciable energy from the pulse. This small terminating impedance makes the input impedance to the $\frac{\lambda}{4}$ line very large.

If an extra $\frac{\lambda}{4}$ line is added to the left of the section, as in

Fig. 179(b), the input impedance changes from an open circuit to a short circuit as the transmitter fires, since the line then acts as a $\frac{\lambda}{2}$ section, (1:1 transformer).

It is an advantage to add to the section an extra $\frac{\lambda}{4}$ line, short-circuited at the termination, as shown in Fig. 179(c). This ensures

that before the valve strikes it is positioned at a voltage antinode and is thus more readily ignited by the transmitted pulse.

This arrangement is incorporated as a shunt element in the parallel circuit shown in Fig. 180(a) and in the series arrangement of Fig. 180(b). Alternatively a similar arrangement may be used as a series element, as in Figs. 180(c) and (d). It is left to the reader to verify that these arrangements satisfy the basic requirements for Fig. 178, described at the beginning of this section.

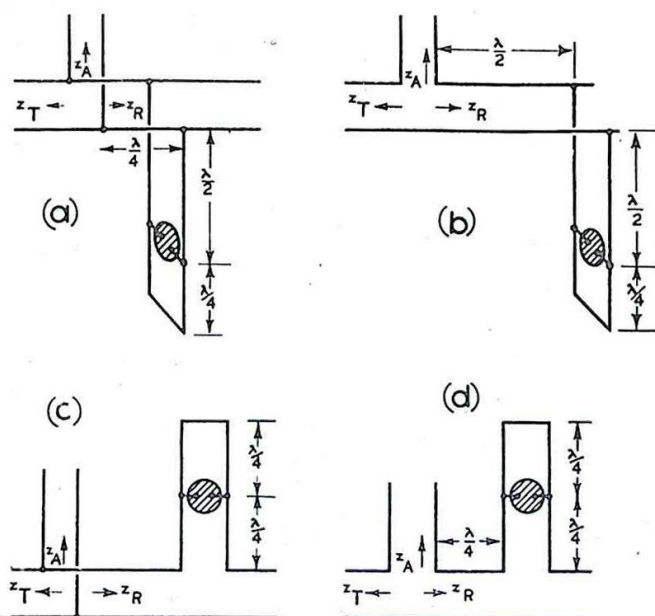


Fig. 180 -
Incorporation of
switching valve in
common T/R circuits.

It should be noted that the $\frac{\lambda}{4}$ or $\frac{\lambda}{2}$ sections fulfil their functions in the steady states only. There is always a build-up time required, during which the wavefront of the transmitter pulse divides at the junction and part is carried to the aerial, part to the switching valve section in the receiver lead, and part is reflected because of the temporary mismatch. This transient period will occupy a number of cycles, but forms only a negligible fraction of the pulse width, provided a sufficiently high frequency is employed.

It is common to use more than one valve or other switching device, at different points on the receiver feeder, as it is more reliable to use several relatively simple switches rather than to attempt to provide a sufficient degree of reliability with a single switch.

Where it is not possible to rely upon the change in output impedance of the transmitter to provide the necessary switching in that branch of the circuit, an additional switching circuit is necessary. Alternative arrangements are shown in Fig. 181. In the parallel case (a) or (c), the input impedance z_T changes from an open circuit, its value when the transmitter is not pulsing and the valve is open, to R_0 when the transmitter fires and closes the valve. In the series case (b) or (d) z_T changes from a short-circuit to R_0 as the transmitter fires.

Neon valves and open or enclosed spark gaps may be used as switching valves with open-wire lines. Open spark gaps may or may not be "blown", and a "keep-alive" electrode, although sometimes desirable, is not always necessary (See Chapter 6).

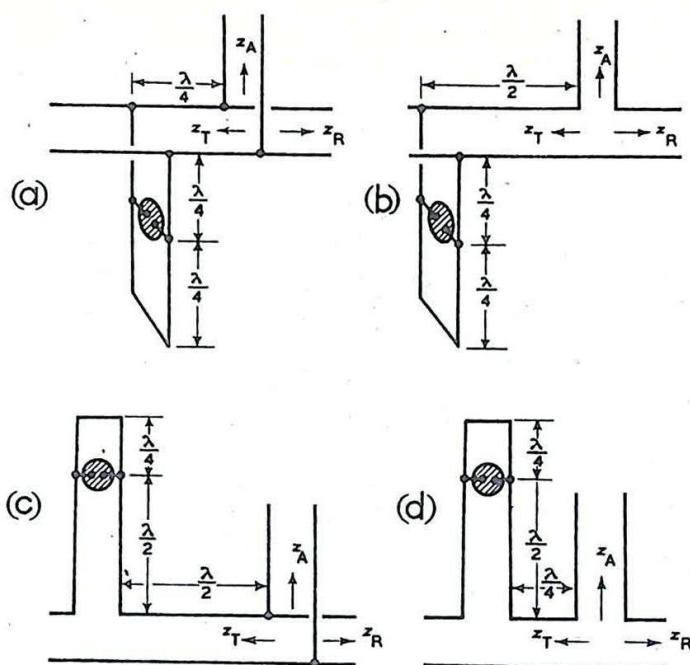


Fig. 181 - Alternative arrangements for switching the transmitter branch of the feeder system.

In the adaptation of this technique to coaxial lines or wave guides a soft rhumbatron is commonly used as a switching device. The valve may be inserted in series with the receiver lead, as shown in Fig. 182(a); the equivalent circuit is shown at (b). The input impedance of this valve is normally resistive and of magnitude R_0 , the rhumbatron acting as a transformer which is adjusted by the positioning of the current loops. When the transmitter fires, the gases in the valve ionize and this condition corresponds to short-circuiting the secondary circuit between A and B. The input impedance becomes small and almost purely reactive. It is this change in input impedance that fulfils the function of the switching valve.

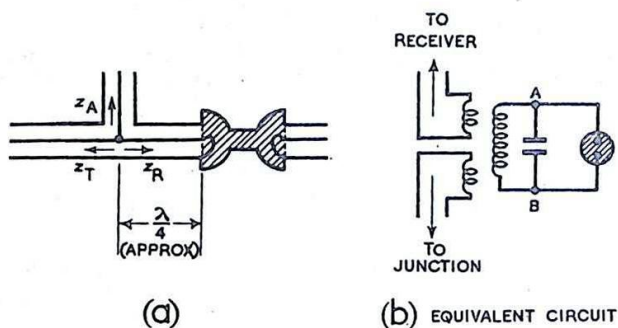
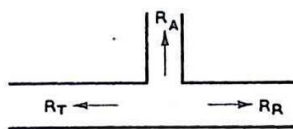


Fig. 182 - Use of soft rhumbatron as switching valve.

Fig. 183 - Simple but inefficient system which relies on change in transmitter impedance.



$$R_T = p R_T \text{ WHEN THE TRANSMITTER IS PULSING.}$$

$$R_T = q R_T \text{ WHEN THE TRANSMITTER IS QUIESCENT.}$$

A very simple but inefficient common T/R system can operate without any switching valves, relying entirely on the change in output impedance of the transmitter and allowing considerable mismatching on reception. This may be illustrated with reference to Fig. 183.

If P_{RT} is the input resistance of the transmitter branch of the feeder system at the junction when the transmitter is pulsing, and Q_{RT} when it is quiescent, the following relations should be satisfied.

- (i) $P_{RT} = R_A + R_R$, and $R_A \gg R_R$; this ensures that on transmission, most of the energy goes to the aerial, and that the system is then properly matched.
- (ii) $R_R \gg Q_{RT}$; this ensures that on reception most of the energy goes to the receiver. However, the previous requirement, $R_A \gg R_R$, implies that a mismatch on receiving is inevitable.

The overall requirements are:-

$$P_{RT} > R_A \gg R_R \gg Q_{RT}.$$

The corresponding relations for the parallel circuit of Fig. 178 are obtained by reversing the inequalities throughout; viz;

$$P_{RT} < R_A \ll R_R \ll Q_{RT}.$$

LIMITATIONS OF TRANSMISSION LINES

41. Resistive and Dielectric Losses

It was stated in Sec. 18 that the propagation constant per unit length of a transmission line may be written

$$\begin{aligned} \gamma_l &= \sqrt{(R_l + j\omega L_l)(G_l + j\omega C_l)} \\ &= \alpha_l + j\beta_l, \end{aligned}$$

where α_l is the loss in nepers per unit length. In Sec. 22 it was further stated that when $\frac{\omega L_l}{R_l} \gg 1$ and $\frac{\omega C_l}{G_l} \gg 1$ it is usually sufficiently accurate to take the loss in nepers as

$$\begin{aligned} \alpha_l &= \frac{1}{2} \left\{ \frac{R_l}{R_0} + G_l R_0 \right\} \\ &= \alpha_c + \alpha_d, \end{aligned}$$

where $\alpha_c = \frac{R_l}{2R_0}$, and is the loss attributable to the finite conductivity of the conductors, and $\alpha_d = \frac{G_l R_0}{2}$, representing the dielectric loss factor.

42. Resistive Losses

For a parallel pair, of radius r and separation d , the value of R_l may be written as

$$R_l \doteq \frac{2}{r} \sqrt{\frac{\mu f}{\sigma}}$$

provided $d \gg r$,

μ being the permeability and σ the conductivity of the metal, and f the frequency.

For a coaxial cable of inner radius r_1 and outer r_2 ,

$$R_l \doteq \sqrt{f} \left\{ \frac{1}{r_1} \sqrt{\frac{\mu_1}{\sigma_1}} + \frac{1}{r_2} \sqrt{\frac{\mu_2}{\sigma_2}} \right\}, \text{ where the suffixes 1}$$

and 2 correspond to the inner and outer conductors respectively.

Using the formula for R_0 given in Section 18, we may write:-

$$\begin{aligned} \alpha_c &= \frac{R_l}{2R_0} \\ &= \frac{\sqrt{f} \left(\frac{1}{r_1} \sqrt{\frac{\mu_1}{\sigma_1}} + \frac{1}{r_2} \sqrt{\frac{\mu_2}{\sigma_2}} \right)}{\frac{138}{\sqrt{K}} \log_{10} \left(\frac{r_2}{r_1} \right)} \end{aligned}$$

In the particular case where $\mu_1 = \mu_2, \sigma_1 = \sigma_2$,

$$\alpha_c = \frac{1}{138} \sqrt{\frac{K\mu f}{\sigma}} \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \log_{10} \left(\frac{r_2}{r_1} \right).$$

From these results we may conclude that for both balanced-pair and coaxial lines:-

- (i) $\alpha_c \propto \sqrt{f}$.
- (ii) for a given R_0 (i.e., a fixed ratio $\frac{r_2}{r_1}$ or $\frac{d}{r}$), α_c is reduced by increasing the size of the conductors.
- (iii) α_c is reduced by using metals with high conductivity and low permeability.

Further results applicable to coaxial cables only are:-

- (iv) There is an optimum value of $\frac{r_2}{r_1}$ which depends on the conditions imposed on the variation of r_1 and r_2 .

Viz. For minimum resistive losses with a constant value for r_2 , $\frac{r_2}{r_1}$ should be 3.6.

For maximum dielectric strength, giving the least chance of a breakdown, $\frac{r_2}{r_1}$ should be 2.72.

- (v) When the conductivities do not have to be equal, the higher conductivity should be given to the inner conductor.

Results applicable to balanced-pair lines only are:-

- (vi) For minimum resistive losses with a constant value for d , $\frac{d}{r}$ should be approximately 4.

For maximum dielectric strength the optimum ratio is approximately $\frac{d}{r} = 5.4$.

43. Dielectric Losses

The power factor F of a dielectric is equal to $\cos \phi$, where $\tan \phi = \frac{\omega C_l}{G_l}$. Since we are concerned with low losses only, we may take

$$\omega C_l \gg G_l$$

$$\text{so that } F \doteq \cot \phi = \frac{G_l}{\omega C_l}.$$

$$\text{Hence } G_l \doteq \omega C_l F.$$

Substituting this in the expression for α_d given in Sec. 41, we have

$$\alpha_d \doteq \frac{\omega C_l}{2} \frac{FR_0}{2} \doteq \frac{\omega F \sqrt{L_l C_l}}{2}, \text{ since } R_0 \doteq \sqrt{\frac{L_l}{C_l}}$$

$$\doteq \frac{2\pi f F}{2u}$$

$$\doteq \frac{\pi f F}{u},$$

where u is the velocity of propagation

in the dielectric.

44. Frequency Effects

Since resistive losses are proportional to \sqrt{F} , and dielectric losses to f , it follows that the former predominate at low, and the latter at high, frequencies. Cables are normally used below the frequency at which dielectric and resistive losses are equal.

45. Coaxial Cables with Low Losses

Cables with the lowest attenuation are those with air as dielectric. The central conductor should be rigidly supported on insulating spacers of low-loss material (e.g., distrene) in order to preserve the characteristic properties of the cable. Such a cable is usually inflexible. Flexible cables require some dielectric filling to support the inner conductor and to keep it central, and a dielectric should be employed, if one exists, such that α_d is less than α_c at the operating frequency.

The radius r_1 , of the inner conductor should be chosen as large as possible consistent with flexibility, and the outer should be given a radius $r_2 = 3.6r_1$, (approximately).

For a solid copper inner conductor a typical value of the diameter is 0.056 inches, with 0.33 inches for the outer. With dielectric constant 2.3, the characteristic impedance would be $R_0 = 75$ ohms. To increase the effective radius of the inner while retaining flexibility, a stranded inner is sometimes employed. This procedure is successful at the lower frequencies, but at higher frequencies stranding increases the loss in the inner. It is important to protect the outer conductor, which is often braided for flexibility, from corrosion, which increases the loss and causes the cable to exhibit inconsistent electrical behaviour. Protection is afforded usually by an outer sheath of polyvinyl chloride or by a coating of enamel. Where a braided outer conductor is used the paths of the currents along the strands of the braid are oblique to the axis, and are effectively lengthened so that the loss in the outer is somewhat increased.

46. Other Losses in Open-wire Feeders

The effect of weather on open wire lines may be serious if the line is several wavelengths long. Damp on insulators and spacers may considerably alter line characteristics and cause standing waves to develop. It is possible to minimise these effects by using conductors stretched between fixed end supports rather than supported at intervals along the line; and these supports should be of the metal insulator type.

Losses occur at frequencies at which the spacing between the conductors is comparable with the wavelength. If there are discontinuities in the line (and at either the beginning or the end some form of discontinuity is inevitable) radiation losses occur. But in any case, the existence of comparatively large induction fields at points near the line may cause considerable loss due to pickup in neighbouring conductors. These conductors may dissipate the energy in the form of heat or, which is usually even more undesirable, radiate electromagnetic energy which would interfere with the directivity of the main aerial system.

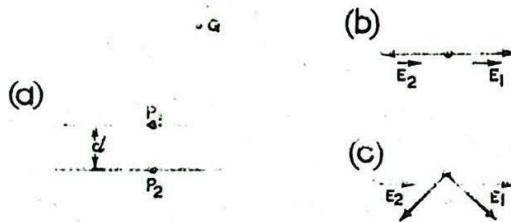


Fig. 184 - Resultant fields due to alternating currents when d is comparable with λ .

The manner in which these fields occur is shown in Fig. 184. The vector \vec{E}_1 represents the time-variable electric field at Q due to the current in a small element of the line at P1, and \vec{E}_2 that due to the corresponding element at P2. These currents are in opposition, so that, provided Q is equidistant from P1 and P2, the equal vectors \vec{E}_1 and \vec{E}_2 will be antiphase and the resultant vector will be of negligible magnitude.

This will be true for all positions of Q provided P1 and P2 are separated by $d \ll \lambda$, as for the case (b). Where d is not small compared with λ , there will be positions of Q for which the difference in pathlengths P1 Q and P2 Q causes an appreciable phase difference in \vec{E}_1 and \vec{E}_2 at Q as at (c). There is thus a resultant vector, whose magnitude is not negligible, at points several wavelengths from the conductors. This normally represents the induction field, but at discontinuities in the conductors the same applies to the radiation field; [for a consideration of these different fields see BR 229 Sec. K para. 7 or AP 1093 Chap. VII para. 31.]

In general, more energy is stored in the electric and magnetic fields at regions further from the conductors and more extraneous radiation of energy occurs as the frequency is raised.

A short length of transmission line improperly terminated may be used as an aerial because of this fact. Some microwave oscillators have open-circuited lines built into the valve-circuits inside the envelope and these fulfil the dual role of resonant circuits and radiators. They may be inserted in waveguides without external connections, energy being radiated direct from the standing wave system on the open-circuited line.

CIRCLE DIAGRAMS

47. Introduction

In Sec. 16 an analytical expression was given for the input impedance of a uniform lossless line of characteristic impedance R_0 terminated in any impedance z_r . Practical problems requiring the use of such relationships in successive applications give rise to complicated arithmetical manipulations which make the analytical method of solution tedious and which mask the physical significance of the processes employed. A geometrical method of tackling such problems involving the use of Circle Diagrams gives results which are sufficiently accurate for most practical work, and are speedily obtainable with a little practice and familiarity with the method. Also the pictorial representation involved in this method helps to keep the physical principles in mind, since movement from one part of a line to another is represented by a particular type of movement from one part of the diagram to another. The subsequent sections are devoted to a description of the Circle Diagram or Transmission Line Calculator and the methods by which it may be used in the solution of transmission line (and waveguide) problems. By this means some of the problems already dealt with qualitatively can be given a quantitative interpretation.

We make no attempt to discuss the theory of circle diagrams but merely show how they can be employed to solve transmission line problems. Further, although two forms of circle diagrams are in common use, the Cartesian and the Polar, we shall here limit ourselves to a description of the former alone.

48. Normalised Impedances and Admittances

The fundamental formula of Section 16,

$$z_s = R_0 \frac{z_r + j R_0 \tan \phi}{R_0 + j z_r \tan \phi}, \quad (\phi = \frac{2\pi l}{\lambda})$$

may be written in the form:

$$\frac{z_s}{R_0} = \frac{\frac{z_r}{R_0} + j \tan \phi}{1 + j \frac{z_r}{R_0} \tan \phi}.$$

If we now replace $\frac{z_s}{R_0}$ and $\frac{z_r}{R_0}$ by $\boxed{z_s}$ and $\boxed{z_r}$ respectively, we have

$$\boxed{z_s} = \frac{\boxed{z_r} + j \tan \phi}{1 + j \boxed{z_r} \tan \phi},$$

so that the formula connecting $\boxed{z_s}$ and $\boxed{z_r}$ depends only on $\frac{l}{\lambda}$ and not on R_0 .

The ratio $\boxed{z} = \frac{z}{R_0}$ is called a Normalised Impedance, and is dimensionless.

Similarly, $\boxed{y} = \frac{y}{G_0}$ is called a Normalised Admittance;

$$(G_0 = \frac{1}{R_0}).$$

* In diagrams, italics are used to denote normalised impedances and admittances.

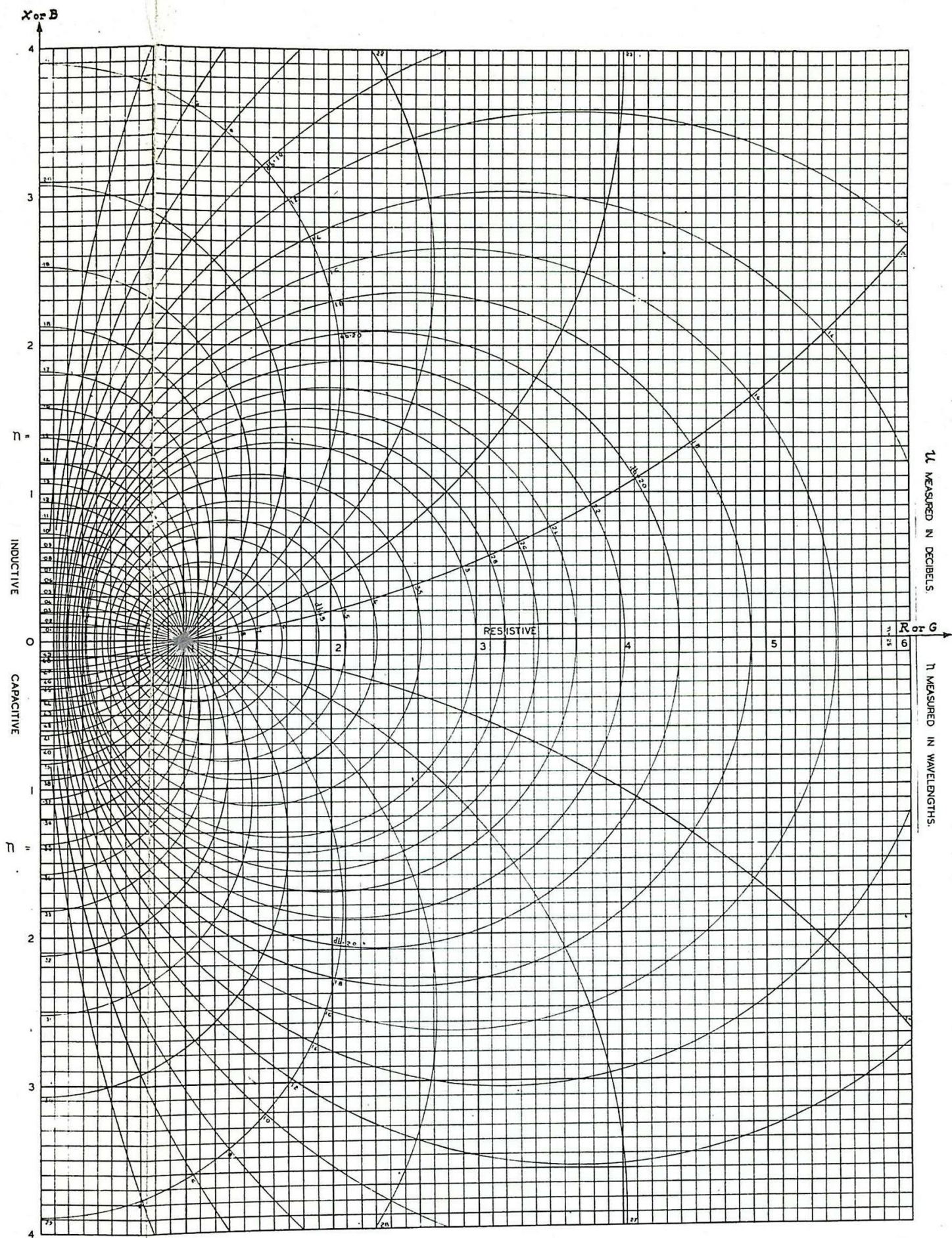


fig. 185. Circle diagram.

Clearly,

$$\begin{aligned}\boxed{y} &= \frac{y}{G_0} \\ &= \frac{R_0}{z} \\ &= \frac{1}{\boxed{z}}\end{aligned}$$

so that the reciprocal relation between admittance and impedance holds also for their normalised equivalences.

The use of normalised impedances and admittances makes it possible to use the same system of co-ordinates for all circle diagrams applied to any uniform loss-less line. Also, the same

normalised quantity \boxed{z} or \boxed{y} may be taken to represent either an impedance or an admittance (normalised) since both are dimensionless.

If $z = R + jX$,

then $\boxed{z} = \boxed{R} + j \boxed{X}$, where \boxed{R} and \boxed{X} are the normalised resistance and reactance respectively, given by

$$\boxed{R} = \frac{R}{R_0} \quad \text{and} \quad \boxed{X} = \frac{X}{R_0}.$$

Similarly, if $y = G + jB$,

$$\text{then } \boxed{y} = \boxed{G} + j \boxed{B}, \text{ where } \boxed{G} = \frac{G}{G_0} \text{ and } \boxed{B} = \frac{B}{G_0}.$$

49. The Cartesian Circle Diagram

Fig. 185 is an example of a Cartesian Circle Diagram. It comprises a Cartesian system of axes as a background which is indicated separately in Fig. 186(a). When working with circle diagrams, impedances and admittances must first be reduced to their normalised forms

$$\boxed{z} = z/R_0 \quad \text{and} \quad \boxed{y} = y/G_0.$$

An impedance $\boxed{z_1} = \boxed{R_1} + j\boxed{X_1}$ is represented on the diagram by the point P $\boxed{R_1}$, $\boxed{X_1}$ whose co-ordinates are $\boxed{R_1}$ and $\boxed{X_1}$ as shown.

The impedance \boxed{z} can be regarded as being represented either by the point P or by the vector \vec{OP} . In the same way an admittance $\boxed{y_1} = \boxed{G_1} + j\boxed{B_1}$ is represented by the vector \vec{OQ} or by the point Q (Fig. 186(a)).

Superimposed on the Cartesian reference system is a family of complete circles, the u-circles, and another family of circular arcs, the n-arcs, that cut the u-circles at right angles. It is not necessary to understand the theory and construction of the diagram in order to use it, but for interest some properties of the

circles are given in the following (starred) paragraph.

Properties of the Circles and Arcs

u-circles are the loci traced out by the point $[z_s]$ given by

$$[z_s] = \frac{[z_r] + j \tan 2\pi n}{1 + j [z_r] \tan 2\pi n}$$

$$-- (n = \frac{f}{\lambda})$$

when $[z_r]$ is kept constant, but n is allowed to vary.

If n is kept constant and $[z_r]$ is allowed to assume various purely resistive values, the locus of $[z_s]$ is an n-arc.

u-circles

It may be shown that the equation of a u-circle in the $[R]$, $[X]$ system of coordinates is

$$([R] - \coth 2u)^2 + [X]^2 = \operatorname{cosech}^2 2u \dots \dots \dots (1).$$

The centre of this circle lies on the $[R]$ axis at the point

$$(\coth 2u, 0) \dots \dots \dots (2),$$

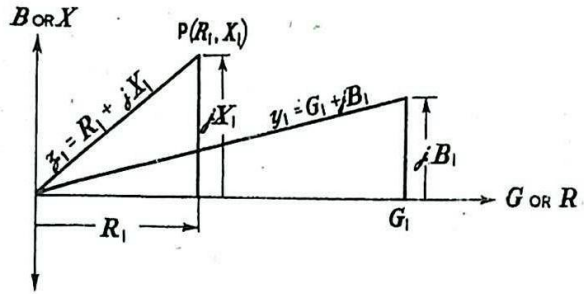
and its radius is,

$$a = \operatorname{cosech} 2u.$$

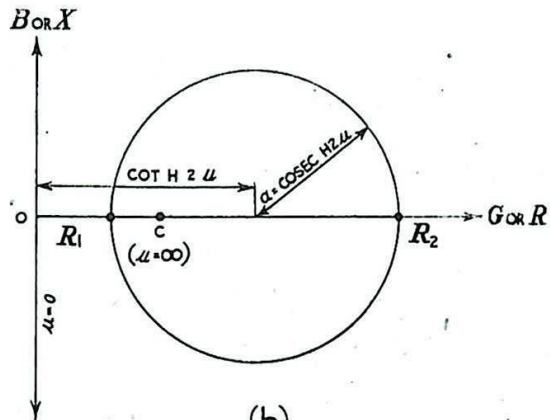
A u-circle is shown in Fig. 186(b).

By assigning to u a sequence of values a family of circles is obtained in which one circle corresponds to one value of u .

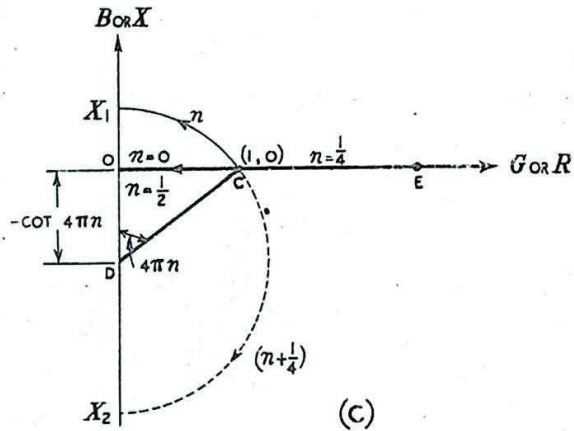
When $u = 0$, equation (2) gives the circle an infinite radius and places its centre at plus infinity along the $[R]$ -axis.



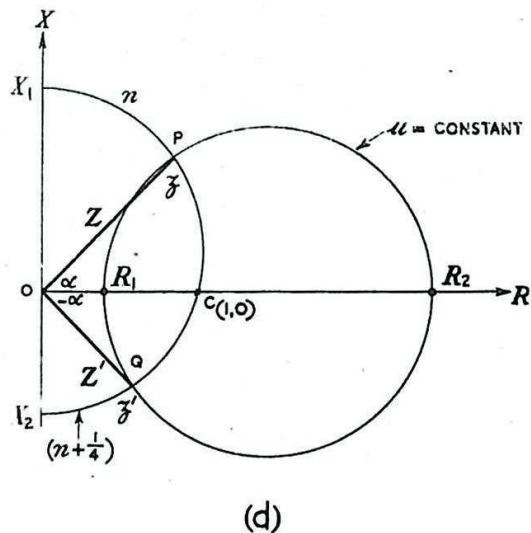
(a)



(b)



(c)



(d)

FIG. 186 - Properties of circle diagram.

The circle $u = 0$ is therefore the same as the $\pm j \boxed{X}$ ($\pm j \boxed{B}$) axis.

At the other extreme, when $u = \infty$, then, according to (2) the radius of the circle is zero and its centre is the point (1,0).

The circle given by

$$u = \infty$$

is the point C in Fig. 186(b). Finite values of u give circles of the type shown.

In practice, the quantity u is associated with the loss in amplitude on reflection at an impedance, or with the attenuation of a travelling wave, and it is convenient to attach to each circle its u -value in decibels instead of its immediate u -value in nepers. This has been done in Fig. 185.

If \boxed{X} is put equal to zero in equation (1) we find that the circle cuts the \boxed{R} -axis at distances $\boxed{R_1}$ and $\boxed{R_2}$ from the origin such that,

$$\boxed{R} = \coth 2u \pm \operatorname{cosech} 2u$$

whence, $\boxed{R_1} = \tanh u$ and $\boxed{R_2} = \coth u$

and $\boxed{R_1} \boxed{R_2} = 1$.

This is equivalent to the equation

$$z_s = R_o^2 / z_r \text{ of section 34, where } z_s \text{ and } z_r \text{ are both resistive.}$$

n-arcs

These are arcs of circles whose equations are,

$$\boxed{R}^2 + (\boxed{X} + \cot 4\pi n)^2 = \operatorname{cosec}^2 4\pi n \dots\dots(3).$$

Their centres therefore lie on the $\pm j \boxed{X}$ axis at the points,

$$(0, -\cot 4\pi n) \dots\dots\dots(4)$$

and their radii are: $\operatorname{cosec} 4\pi n$.

For each value of n within the range 0 to $\frac{1}{2}$ we obtain an arc, but the sequence of arcs repeats if n is increased beyond the value of $\frac{1}{2}$.

An n -arc is shown in Fig. 186(c). According to equation (4) and Fig. 186(c), since in the triangle ODC, $OD = -\cot 4\pi n$ and $DC = \operatorname{cosec} 4\pi n$, the angle $ODC = -4\pi n$ and $OC = 1$.

Thus, all n -arcs start out from the unit point C (1,0).

When $n = 0$ the centre of the arc lies at minus infinity on the $j \boxed{X}$ axis. The arc $n = 0$ is therefore the portion CO of the real axis. As n is increased the centres D move up the $\pm j \boxed{X}$

axis and finally reach the origin when $n = 1/8$. As n increases further the centre D continues to move upwards and disappears to plus infinity at $n = 1/4$.

The portion CG of the real axis is the arc $n = 1/4$. When n increases in a series of equal increments a set of arcs is obtained in the upper half of the diagram as shown. Those in the lower half are easily obtained by completing the semicircle of which each n -arc in the upper half is a portion. The completing arc, dotted in Fig. 186(c), is called the complementary arc. If the n -value of an upper arc is n , then that of its complementary arc is $(n + 1/4)$. Thus, the n -values of the upper arcs range from $n = 0$ to $n = 1/4$, and of the lower arcs from $n = 1/4$ to $n = 1/2$. The arc $n = 1/2$ is again the line CO which is also the arc $n = 0$.

If n exceeds the value of $1/2$ and is for instance equal to $n = (\frac{h}{2} + n')$ where $h = 1, 2, 3$, etc. then the arc n is the same as the arc n' .

The distances from the origin $\boxed{X_1}$ and $\boxed{X_2}$ at which the same semicircle (arcs n and $n + 1/4$) cuts the $\pm j\boxed{X}$ axis are obtained by putting $\boxed{R} = 0$ in equation (3).

$$\text{Thus, } \boxed{X} = \pm \operatorname{cosec} 4\pi n - \cot 4\pi n = \pm \frac{1 - \cos 4\pi n}{\sin 4\pi n}$$

That is,

$$\boxed{X_1} = \tan 2\pi n; \quad \boxed{X_2} = -\cot 2\pi n \quad \dots\dots\dots(5)$$

as shown in Fig. 81(c).

It follows that,

$$\boxed{X_1} \boxed{X_2} = -1 \quad \dots\dots\dots(6).$$

If $\boxed{X_1}$ is a normalised reactance then $\boxed{X_2}$ is the corresponding normalised susceptance and conversely.

50. Applications of the Circle Diagram

Each u -circle corresponds to a particular standing wave distribution on the line. All points corresponding to input impedances at various positions on the same standing wave system lie on the same u -circle. The application of circle diagrams to lossy lines, in which the standing wave ratio varies from point to point, is discussed in Sec. 53.

Each n -arc corresponds to the position of points relative to the standing wave system in which they are located. The difference between two values of n corresponds to the distance along the line, measured in wavelengths; i.e.

$$n_1 - n_2 = \frac{l}{\lambda}$$

An important property of the diagram is illustrated in Fig. 186(d). It shows a complete u -circle enclosing the unit point $C(1, 0)$ and two n -arcs, CP and CQ that intersect it at P and Q. The n -value of each n -arc is marked on it near its end and each u -circle

also is marked with its u -value (Fig. 185). As an inspection of the chart will reveal the values of n run from zero for the arc CO to $n = 1/2$ as CO is reached again from below, whereas the values of u range from zero for the reactive axis of co-ordinates to infinity at the limiting point C . In Fig. 186(d) the arcs CP and CQ belong to the same semicircle and are called complementary. It will be found that if the n -value of any arc CP is n , then that of its complementary arc CQ is $(n + \frac{1}{2})$.

Further, if OP represents the normalised impedance $z = \boxed{R} + j\boxed{X}$ (i.e. P is the point \boxed{R} , \boxed{X}) then OQ represents the normalised impedance (or normalised admittance) $\boxed{z'} = 1/\boxed{z}$. By this we mean that if the length of OP is \boxed{Z} , the magnitude of \boxed{z} , and the angle \widehat{COP} is α , then the length of OQ is $\boxed{Z'} = \frac{1}{\boxed{Z}}$

and the angle COQ is $\alpha' = -\alpha$. It must be stressed that for this to be true CP and CQ must be complementary arcs.

Thus, if $\boxed{z} = \boxed{R} + j\boxed{X}$ is a normalised impedance represented by \overrightarrow{OP} , then the associated normalised admittance $\boxed{y} = \frac{1}{\boxed{z}} = \boxed{G} + j\boxed{B}$ is represented by \overrightarrow{OQ} . (Fig. 186(d)).

In particular, if the angle α is zero then P lies on the real axis and \overrightarrow{OP} represents a pure normalised resistance, $\boxed{R_2}$ say. The point Q will also fall on the real axis at the opposite end of the diameter and will correspond to the conductance $\boxed{G_1} = \boxed{R_1}$, such that $\boxed{R_1}\boxed{R_2} = 1$.

Similarly, when $\alpha = 90^\circ$ the arc CP meets the imaginary axis in the reactance point $+j\boxed{X_1}$. Then the arc CQ meets this axis at the corresponding susceptance point $-j\boxed{X_2}$ such that $\boxed{X_1}\boxed{X_2} = -1$.

51. Numerical Examples

To find the Input Impedance z_s given R_o , z_r , l and λ . The line is assumed to be loss-free. Refer to Fig. 187(a).

Procedure

1. First normalise z_r , the terminating impedance; thus

$$z_r = \frac{z_r}{R_o} = \frac{R_r}{R_o} + j\frac{X_r}{R_o} = \boxed{R_r} + j\boxed{X_r}.$$

2. Plot the point $\boxed{z_r} = \boxed{R_r} + j\boxed{X_r}$ on the cartesian diagram (Point P , Fig. 187(a)).

3. Note the value of

- (i) the n -arc n_r ,
- (ii) the u -circle u_r .

that pass through P (or estimate these values by interpolation between circles and arcs).

4. Evaluate ℓ/λ and add it to n_r to obtain

$$n_s = n_r + \ell/\lambda.$$

Since the line is loss-free, $u_s = u_r$.

5. Traverse in a clockwise sense the u -circle $u = u_r$ until it meets the n -arc $n = n_s = n_r + \ell/\lambda$, at the point Q. At this point the normalised impedance is z_s :-

$$\boxed{z_s} = \boxed{R_s} + j \boxed{X_s}.$$

The input impedance is, as required

$$z_s = R_0 \boxed{R_s} + j R_0 \boxed{X_s} = R_s + j X_s.$$

Example 1

A loss-free transmission line whose characteristic impedance is 300 ohms and whose length is 0.3λ is terminated by a load, $z_r = (600 + j 300)$.

Find the input impedance.

The normalised load is

$$\boxed{z_r} = \frac{z_r}{R_0} = (2 + j1).$$

From Fig. 185 the u and n -values of the impedance point $\boxed{z_r} = 2 + j1$, are:-

$$u_r = 3.5 \text{ db}; \quad n_r = 0.213.$$

Evaluate

$$u_s = u_r; \quad n_s = 0.213 + \ell/\lambda = 0.513.$$

The arc $n = 0.513$ is the same as the arc $n_s = 0.013$.

The point of intersection of the circle $u = u_r = 3.5 \text{ db}$ and the arc $n = n_s = 0.013$ is

$$\boxed{z_s} = (0.38 + j 0.07).$$

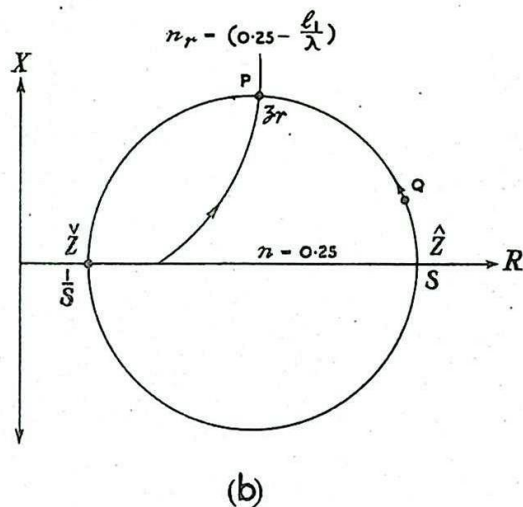
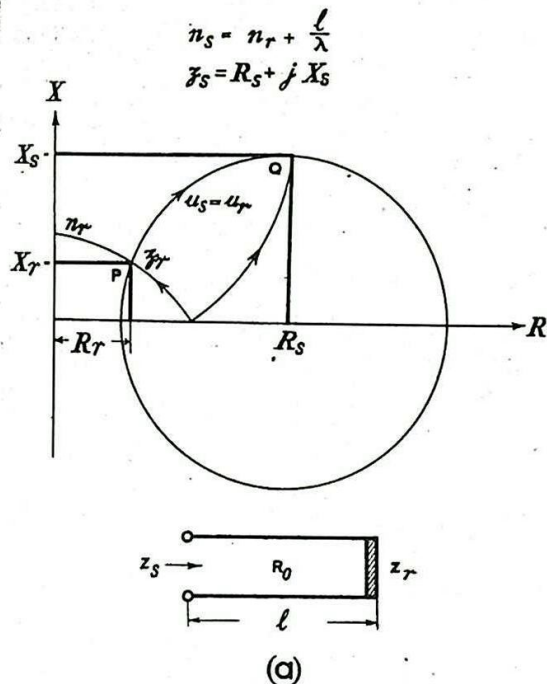


fig. 187 - Use of circle diagrams.

The input impedance is therefore

$$z_s = 300 \boxed{z_s} = (114 + j.21).$$

Example 2

The transmission line of example 1 is terminated by an admittance $y_r = G_r + j B_r = (0.01 + j 0.005)$.

Find the input admittance.

We treat admittances exactly as if they were impedances. We therefore follow the procedure formulated above:-

The normalised admittance is,

$$\boxed{y_r} = \frac{y_r}{G_0} = R_0 y_r = 300 y_r = (3 + j 1.5).$$

Plot the point $\boxed{y_r} = (3 + j 1.5)$ and note u_r and n_r . From Fig. 185 we read,

$$u_r = 2.32 \text{ db}; \quad n_r = 0.227.$$

$$\text{Form } n_s = n_r + l/\lambda = 0.227 + 0.3 = 0.527.$$

We may therefore take n_s to be 0.027.

The point of intersection of the arc $n_s = 0.027$ and the circle $u_r = 2.32$ is,

$$\boxed{y_s} = 0.28 + j 0.17.$$

The input admittance is,

$$y_s = G_0 \boxed{y_s} = \frac{\boxed{y_s}}{R_0} = \frac{0.28 + j 0.17}{300} \\ = 0.0009 + j 0.00057.$$

Example 3

Find the normalised admittance that corresponds to the impedance $\boxed{z_r}$ of example 1.

According to Sec. 50, and Fig. 186(d) the admittance of the impedance $\boxed{z_r}$ which lies at the intersection of the circle u and arc n (vector \overrightarrow{OP} Fig. 186(d)) is represented by the point of intersection of the circle u and the complementary arc $(n + \frac{1}{2})$.

From example 1,

$$\boxed{z_r} = 2 + j 1; \quad u_r = 3.5 \text{ db}; \quad n_r = 0.213.$$

Consequently, for $\boxed{y_r} = \frac{1}{\boxed{z_r}}$

$$u = u_r = 3.5 \text{ db}; \quad n = (n_r + \frac{1}{2}) = 0.463.$$

From Fig. 185.

$$\boxed{y_r} = (0.4 - j 0.2).$$

(By calculation: $\boxed{y_r} = \frac{1}{\boxed{z_r}} = \frac{1}{2 + j1} = \frac{2 - j1}{5} = 0.4 - j 0.2$)

Example 4

The transmission line of example 1 is terminated by an inductive reactance $X_r = 150$ ohms. Find z_s when $l = 0.3\lambda$.

We have:-

$$\boxed{X_r} = \frac{X_r}{R_0} = \frac{150}{300} = 0.5; \quad \boxed{z_r} = 0 + j. 0.5.$$

From Fig. 185

$$u_r = 0; \quad n_r = 0.074.$$

For $\boxed{z_s}$; $u_s = u_r = 0$; $n_s = n_r + l/\lambda = 0.074 + 0.3 = 0.374$.

From the chart $\boxed{z_s} = -j1$. Hence z_s is a capacitive reactance of 300 ohms.

Example 5

Find the standing wave ratio S on the transmission line of example 1.

According to Sec. 13, $S = \frac{\hat{Z}}{R_0} = \boxed{\hat{Z}}$.

The point $\boxed{z_r} = 2 + j1$ lies on the u -circle $u_r = 3.5$ db and arc $n_r = 0.213$.

The impedance z_s becomes purely resistive at the points $\boxed{\hat{Z}}$ and $\boxed{\check{Z}}$ where this circle cuts the resistive axis.

From Fig. 185, this occurs at

$$\boxed{\hat{Z}} = 2.53; \quad \boxed{\check{Z}} = \frac{1}{2.53} = 0.395.$$

Whence $S = 2.53$.

The distance l_1 of $\boxed{\hat{Z}}$, which is that of the nearest voltage antinode from $\boxed{z_r}$ is given by:

$$\Delta n = 0.25 - 0.213 = 0.037 = l_1/\lambda.$$

Whence $l_1 = 0.037\lambda$.

Similarly the distance of the nearest voltage node is at

$$l_2 = (0.5 - 0.213)\lambda = 0.287\lambda.$$

52. Use of Circle Diagram to Determine the Magnitude of the Load Impedance from a Knowledge of the Standing Wave Pattern

A knowledge of the standing wave pattern enables us to determine S and λ , and shows the distances of the voltage nodes and antinodes from the termination. At voltage antinodes and nodes the line impedance becomes purely resistive and attains its maximum and minimum values \hat{Z} and \check{Z} .

Further (see Sec. 13)

$$S = \frac{\hat{Z}}{R_0} = \frac{R}{\check{Z}}, \text{ so that}$$

$$\hat{Z} \check{Z} = R_0^2.$$

We may write

$$S = \frac{\hat{Z}}{\check{Z}} = \frac{1}{\frac{\check{Z}}{\hat{Z}}}.$$

Since S determines $\frac{\hat{Z}}{\check{Z}}$ and $\frac{\check{Z}}{\hat{Z}}$, the representative u-circle is determined uniquely by the standing wave ratio, $(S, 0)$ and $(\frac{1}{S}, 0)$ being the ends of its resistive-axis diameter.

As the distance l from the load is increased the representative point Q of the impedance $\boxed{z_s}$ (Fig. 187(a)) traverses its u-circle.

A displacement of $\lambda/2$ along the line from any position takes Q exactly once round the u-circle so that it returns to the original impedance $\boxed{z_s}$. Thus the magnitude of the normalised impedance oscillates between the extremes of $\frac{\hat{Z}}{\check{Z}}$ and $\frac{\check{Z}}{\hat{Z}}$ as Q continues to traverse the circumference.

When S , λ and l_1 (the distance of the first voltage antinode from the termination) are known from the standing wave pattern then the procedure of example 5, Sec. 51, may be reversed to give z_r .

Thus we plot $\frac{\hat{Z}}{\check{Z}} = S$ on the real axis (arc $n = 1/4$) as shown in Fig. 187(b).

As we move away from the generator towards the load the moving point Q traverses the u-circle in a counter-clockwise sense.

If, therefore, Q starts at S , the position of voltage antinode, it reaches the point P corresponding to z_r where the u-circle through S cuts the arc $n = n_r = (0.25 - l_1/\lambda)$, since z_r lies at a distance l_1 from the voltage antinode on the side away from the generator. The normalised line impedances at all other positions in this standing wave pattern are represented by points on the u-circle through S .

Example

Suppose $S = 5$ and $l_1 = 0.1\lambda$. Find z_r .

Then $u = 1.8$. The intersection of the circle $u = 1.8$ with

the arc $n = 0.25 - 0.1 = 0.15$ is the point $\boxed{z_r} = 0.55 + j 1.23$.

53. Application of Circle Diagrams to Lines with Low Losses

When the line loss cannot be neglected then it is not only necessary to evaluate $n_s = n_r + \ell/\lambda$ but also to determine the value of u_s which is no longer equal to u_r ; that is, the representative point does not remain on the same u -circle for different lengths of the line. If the movement on the line is away from the load towards the generator then u_s increases with ℓ and Q moves on the diagram in a spiral towards the limiting point C (1,0). When ℓ becomes large then the input impedance z_s is always represented by a point very near C. This is equivalent to saying that the normalised input impedance of a long attenuating transmission line is

$$\boxed{z_s} \doteq 1; \text{ i.e., } z_s \doteq R_0.$$

The conversion of u to u_s is made as follows. It is supposed that the signal loss in decibels in a standard length of the line (say 100 feet) is known. Suppose that the transmission line shown in Fig. 187(a) produces an attenuation in a travelling wave of a_ℓ decibels per unit length and therefore $a = a_\ell \ell$ decibels for length ℓ . To find the normalised input impedance $\boxed{z_s}$ when $\boxed{z_r}$ is known

proceed as follows:-

Locate $\boxed{z_r}$ as before on the diagram and note u_r and n_r .

(u_r in decibels).

Evaluate $u_s = (u_r + a)$ and $n_s = (n_r + \ell/\lambda)$.

Move clockwise around the u_r circle up to the arc n_s . Move inwards along the n_s arc to its point of intersection with the circle u_s .

This determines the representative point of $\boxed{z_s}$.

Example

A resistance of 200 ohms terminates a 50 ft length of Uni-radio 1 cable ($R_0 = 72$ ohms; $a_\ell = 3.5$ db. per 100 feet at 200 Mc/s. and $K \doteq 2.25$). Find the input impedance at a frequency of 200 Mc/s.

$$a = a_\ell \ell = 3.5 \times \frac{50}{100} = 1.75 \text{ db.}$$

The wavelength in air is $1\frac{1}{2}$ metres and in the cable is

$$\frac{3}{2} \cdot \frac{1}{\sqrt{K}} = 1 \text{ metre.} \quad \text{Whence } \ell/\lambda \doteq 15.25.$$

$$\boxed{z_r} = \frac{200}{72} = 2.78 + j.0$$

From Fig. 185

$$u_r = 3.3; \quad n_r = 0.$$

$$u_s = 3.3 + 1.75 = 5.05 \text{ db; } n_s = n_r + 15.25 = 15.25.$$

The arc $n = n_s$ is equivalent to the arc $n = 0.25$.

From Fig. 183

$$\boxed{z_s} = 1.9 + j.0$$

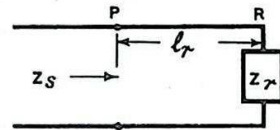
Hence $z_s = 72 \boxed{z_s} = (137 + j.0).$

54. Application of Circle Diagrams to Matching Devices constructed from Sections of Uniform Loss-free Transmission Line

(i) Important Property of Circle Diagrams

Refer to Fig. 188; let \overrightarrow{OR} represent the impedance $\boxed{z_r}$, and \overrightarrow{OP} the corresponding input impedance at the point P distance l_r from R. If $\boxed{z_r}$ varies in such a manner that the point R traces out a

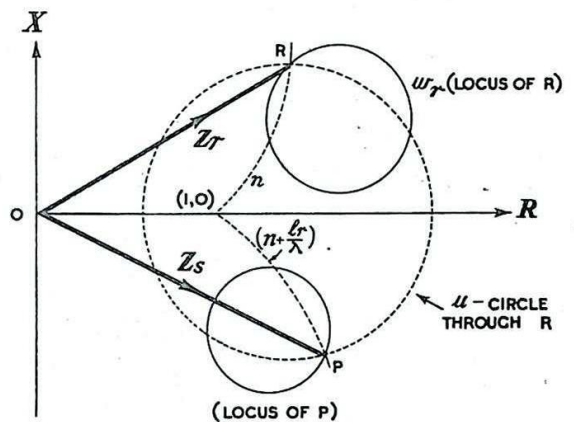
circle, then it can be shown mathematically or by plotting a series of points that the point P also traces a circle, (or, in some cases, a straight line).



(a)

The method of plotting the point P corresponding to a particular position of R is indicated in the figure. In the particular case when the circle-locus of R degenerates into a straight line (circle of infinite radius) the locus of P is still in general, a circle.

The property is of importance in the demonstration of some matching problems using circle diagrams. It is used in the cases given below, of Double Stub Matching and Matching by Slugs.



(b)

*** Mathematically, the relation

Fig. 188 - Important property of circle diagrams.

$$\boxed{z_s} = \frac{1 + j \boxed{z_r} \tan \frac{2\pi l}{\lambda}}{\boxed{z_r} + j \tan \frac{2\pi l}{\lambda}}, \text{ where } l \text{ is constant,}$$

is of the form

$$\boxed{z_s} = \frac{\alpha + \beta \boxed{z_r}}{\gamma + \delta \boxed{z_r}}, \text{ where, in general, } \alpha, \beta, \gamma, \delta \text{ are complex as well as } \boxed{z_s} \text{ and } \boxed{z_r}.$$

This may be written in the form

$$\boxed{z_s} = \alpha' + \frac{\beta'}{\boxed{z_r} + \gamma'}, \text{ and is equivalent to the following steps:-}$$

- (1) a translation,
- (2) an inversion,
- (3) a magnification and a rotation, and
- (4) a further translation.

None of these steps distorts the circular shape of the locus of $\boxed{z_r}$.

[Step (2) may transform a circle into a straight line - i.e., a circle of infinite radius - or vice versa.]

(ii) $\lambda/2$ Transformer

The action of this, the simplest of line transformers, is demonstrated by the movement of the point P (Fig. 186(d)) once completely round the appropriate u-circle.

(iii) $\lambda/4$ Transformer

In this case the point P, corresponding to $\boxed{z_r}$ (Fig. 186(d)) traverses the appropriate u-circle from the n-arc to the $(n + \frac{1}{4})$ -arc, i.e. to the point Q.

(iv) Double $\lambda/4$ Transformers

Consider first the behaviour of a single $\frac{\lambda}{4}$ transformer when a small increase in frequency causes the line to become rather more than $\frac{\lambda}{4}$ in length. This is illust-

trated in Fig. 189(a), for the case in which input and output impedances are required to be resistive. If the terminating impedance is R_r , (assume $R_o > R_r$), normalised impedance is given by $\frac{1}{S} = \frac{R_r}{R_o}$, where S is the standing wave ratio, and is represented by R. The input impedance $\frac{R_o^2}{R}$ when normalised is $\frac{R_o}{R}$ and

is represented by P(S, 0) provided the length of the transformer is exactly $\frac{\lambda}{4}$. If this

is increased a reactive term is introduced, illustrated by the point P'. Although the resistive component of the input impedance is still approximately correct a considerable phase error may be introduced for a small change in wavelength.

Fig. 189(b) shows a double $\frac{\lambda}{4}$ transformer used

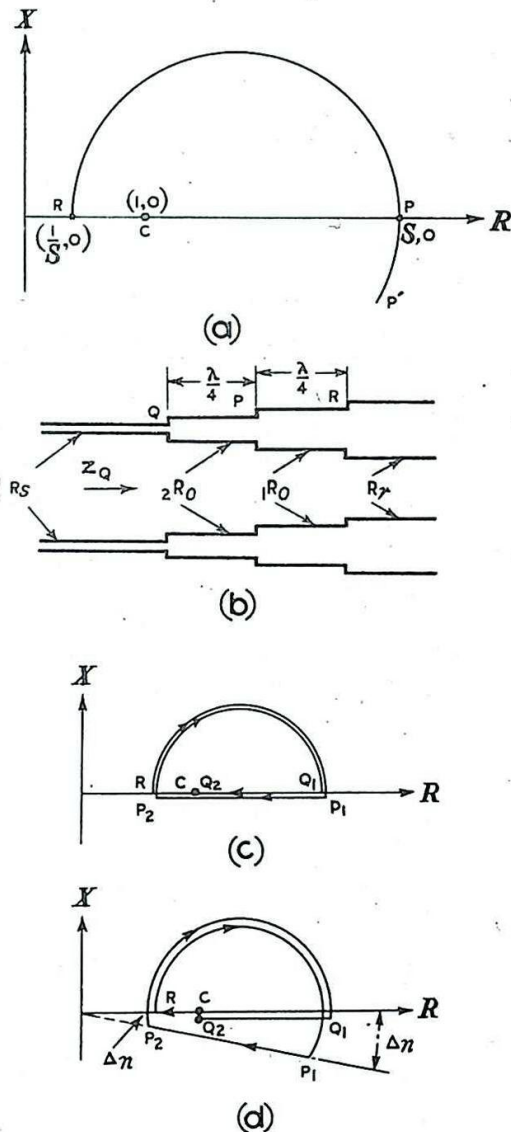


Fig. 189 - Double $\lambda/4$ transformer.

to minimise the above effect. The characteristic impedances, ${}_1R_0$ and ${}_2R_0$ of the $\frac{\lambda}{4}$ sections of line are related to the resistances at the receiving and sending ends of the line, in this case the characteristic resistances of the respective feeders, by the relations

$${}_1R_0 = a R_r \text{ and } {}_2R_0 = a^2 \cdot {}_1R_0, \text{ where } \frac{R_s}{R_r} = a^4.$$

The change of characteristic impedance at the junction generally necessitates changing from one u-circle to another, since for each separate section of line the impedance considered must be normalised with respect to the appropriate characteristic resistance.

This is illustrated at (c) for the case when each section of the double-transformer is exactly $\frac{\lambda}{4}$ in length. Since $\frac{R_r}{{}_1R_0} = \frac{1}{a}$,

the standing wave ratio for the first $\frac{\lambda}{4}$ section is a , and the point R of the circle diagram, corresponding to the input impedance at R (Fig. 189(b)), normalised with respect to ${}_1R_0$, is the point $(\frac{1}{a}, 0)$. The point C (1,0) represents the terminating impedance R_r normalised with respect to itself. Hence we may consider the change in characteristic impedance at R to be represented on the diagram by movement of the representative point from C to R. The change due to the movement from R to P (Fig. 189(h)) is represented by movement from R to P_1 along the u-circle. Since R is the point $(\frac{1}{a}, 0)$, P_1 is the point $(a, 0)$. The change in characteristic impedance from ${}_1R_0$ to ${}_2R_0$ at P necessitates dividing the normalised impedances by a^2 , since ${}_2R_0 = a^2 {}_1R_0$, so that $\frac{z}{{}_2R_0} = \frac{1}{a^2} \cdot \frac{z}{{}_1R_0}$. This transfers the representative point to P_2 , which is $(\frac{1}{a}, 0)$, the same as R. Movement from P to Q (Fig. 189(b)) is represented by movement along the u-circle from P_2 to Q_1 in Fig. 189(c). Finally, since $R_s = a \cdot {}_1R_0$, the change in impedance levels at Q is represented by dividing the new impedance by a ; i.e., the representative point returns to $Q_2 \equiv C$ and the line is properly matched.

If the frequency is slightly increased so that the electrical length $\frac{l}{\lambda}$ of each of the matching sections is increased by the same

amount, the conditions are altered to those shown in Fig. 189(d). Provided the frequency shift is not too large the change in n-value at P_2 is the same as at P_1 (this can be verified from Fig. 185) so that to a first approximation Q_1 is not shifted, and the input impedance represented by Q_2 satisfies the requirements for a broad-band match.

A similar argument shows that any odd number of transformers with appropriate characteristic impedances, used in cascade, is sensitive to changes in frequency, whereas any even number is not. However, the more $\frac{\lambda}{4}$ sections there are inserted between R_r and R_s the

smaller is the standing wave ratio on each section and the more closely do the appropriate u-circles approach the point C. If a large number of such transformers is used, with their characteristic impedances exponentially graded, a wide-band match is achieved irrespective of whether the number of sections is odd or even.

(v) Single Shunt Stub

This matching device is illustrated in Fig. 190(a). The position of the stub, distant ℓ_1 from the termination $T_1 T_2$ is variable, and also the length ℓ_2 of the stub. It is required to adjust ℓ_1 and ℓ_2 so that the input admittance at $P_1 P_2$ has a required value y_s (usually the characteristic admittance of the line connected to $P_1 P_2$).

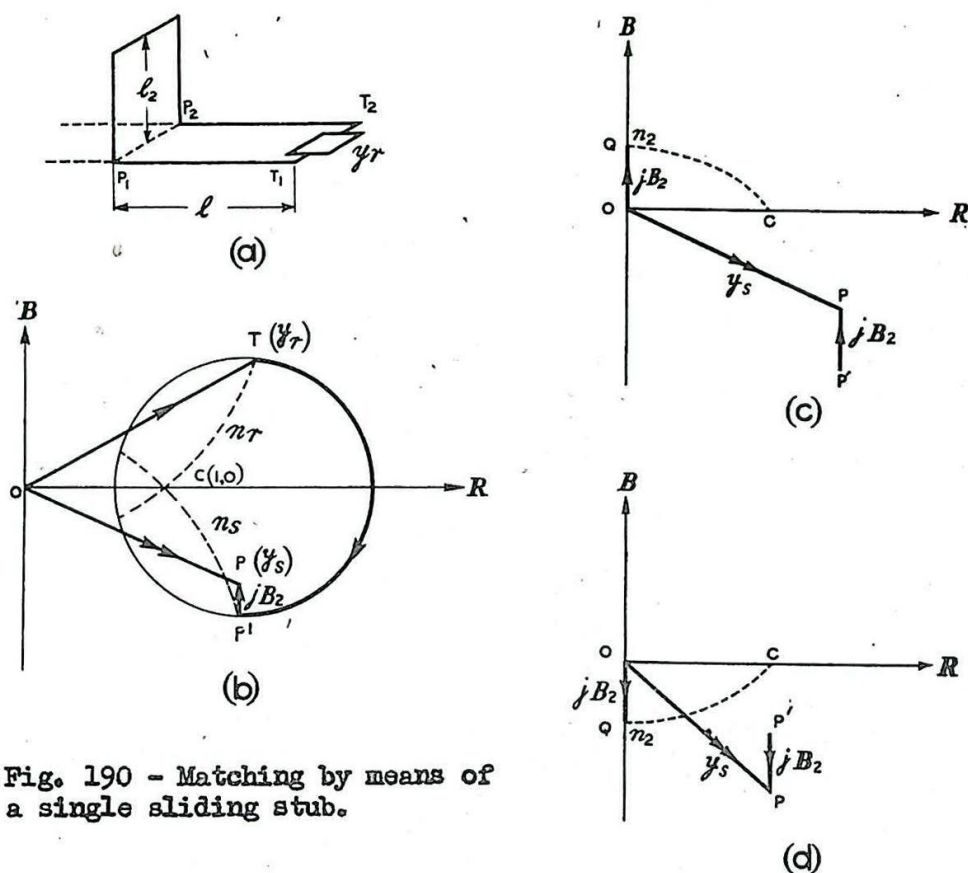


Fig. 190 - Matching by means of a single sliding stub.

The terminating admittance y_r and the characteristic impedance R_0 determine the standing wave pattern and the appropriate u -circle $u = u_r$ for the section PT (Fig. 190(b)). On this circle the points T and P (representing the required input admittance) are known. The stub-length ℓ_2 is adjusted until a susceptance jB_2 is shunted across the line at $P_1 P_2$. This procedure determines the point P' . (There are two alternative positions for P' corresponding to two complementary solutions to the problem. They are the intersections with the circle $u = u_r$ of a vertical line drawn through

y_s . Only one of these two positions is shown on the diagram).

P' determines the value n_s of the required n -arc. The distance ℓ_1 between stub and termination is given by $\frac{\ell_1}{\lambda} = n_s - n_r$.

If y_s may have any value a match is not always possible, i.e. the line through y_s parallel to the susceptance axis may not intersect the circle $u = u_r$. In most cases however y_s is the

characteristic admittance of a line identical with that forming the section PT so that $[y_s] = 1$. In this case $[y_s]$ is the point C (1,0) and two solutions are always possible, one of which corresponds to a length l_1 less than, and the other greater than $\lambda/4$. The shorter length is usually preferable. The length l_2 corresponding to the susceptance $[B_2]$ may be obtained from the circle diagram as indicated below or from the formulae:

$$[B_2] = -\cot \frac{2\pi l_2}{\lambda} \text{ for a short-circuited stub.}$$

$$[B_2] = +\tan \frac{2\pi l_2}{\lambda} \text{ for an open-circuited stub.}$$

The determination from the circle diagram of the length l_2 corresponding to the susceptance $[B_2]$ is illustrated at (c). For an open-circuited stub draw OQ so that $OQ = [B_2]$. The value n_2 of the n-arc through Q then gives l_2 from the relation

$$n_2 = \frac{l_2}{\lambda}$$

It is important that the direction of the admittance $j[B_2]$ is determined correctly. In the case illustrated at (d) the point Q lies below the origin, corresponding to a value of n_2 between 0.25 and 0.5.

If a short-circuited stub is used 0.25λ must be added to, or subtracted from, the length l_2 obtained above for the open-circuited stub.

Numerical example

The results of an actual experiment are quoted in illustration.

The unscreened twin transmission line of Fig. 191 comprised a pair of copper wires in tension. The characteristic impedance was $R_0 = 320$ ohms.

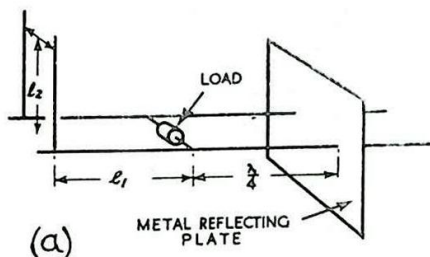
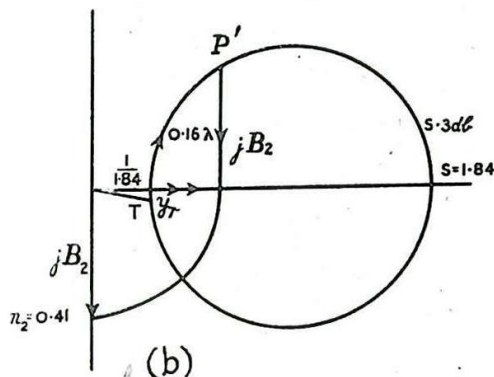


Fig. 191 - Measurement of an impedance at $1\frac{1}{2}$ Mc/s.



The lines with the stub omitted were fed from a low power, loosely coupled 200 Mc/s. oscillator and a complete standing wave was produced on the line by using the metal plate termination shown. The wavelength was obtained from the standing wave pattern using a standing wave indicator.

The measured wavelength was 151 centimetres.

It was found that a voltage antinode existed exactly at $\lambda/4$ from the plate which behaved therefore as a termination of zero impedance.

A resistor whose nominal value was 560 ohms was shunted across the line at $\lambda/4$ from the plate, i.e. where the line impedance was infinite. The terminating impedance of the line to the left was therefore that of the resistor alone. A standing wave ratio $S = 1.84$ was found on the line and a voltage node was located at $\ell_n = 39$ centimetres from the resistor towards the generator. At this point the normalised admittance of the line is $S + j.0 = 1.84 + j.0$ (See Sec. 52). This point lies on the circle $u = 5.3$ db. and the arc $n = 0.25$. We have $\ell_n/\lambda = \frac{39}{151} = 0.258$. Consequently,

the admittance point $\boxed{y_r}$ of the termination is the point of intersection of the circle $u = 5.3$ db. and the arc $u = (0.25 - 0.258) = -0.008$ which is equivalent to the arc $n = 0.492$. This gives $\boxed{y_r} = 0.54 - 0.04j$

$$\text{or } \boxed{z_r} = \frac{0.54 + 0.04j}{(0.54)^2 + (0.04)^2} = 1.8 + j. 0.136.$$

$$z_r = 320 \boxed{z_r} = (576 + j. 4.36),$$

i.e. the "560 ohm resistor" has in fact a resistance of 576 ohms in series with an inductive reactance of 4.36 ohms.

The standing wave on the line was eliminated by the use of a short-circuited shunt stub as in Fig. 191(a). The point y_r occupies a position on the circle $u = 5.3$ db. as shown in Fig. 191 (b). The length of line from y_r to P' is given by

$$\Delta_n = 0.15 - 0.49 = -0.34 \text{ which is equivalent to}$$

$$\Delta_n = 0.16$$

Hence the stub must be placed 0.16λ from the load, i.e.
 $\ell_1 = 24.2$ cms.

Since a short-circuited stub is used its length ℓ_2 is given by $\ell_2/\lambda = 0.25 + n_2$. From the circle diagram we obtain $n_2 = 0.41$, so that $\ell_2/\lambda = 0.66$ which is equivalent to $\ell_2/\lambda = 0.16$. Hence ℓ_2 also is 24.2 cms.

(vi) Double Shunt Stub

Fig. 192(a) illustrates the use of two shunt stubs of variable lengths ℓ_1 and ℓ_2 separated by a fixed distance ℓ_2 for matching a line

to a terminating admittance y_r . It will be assumed, as is common in practice, that all portions of the line and stub have the same characteristic admittance G_0 . It is thus required to adjust the lengths l_1 and l_2 so that the terminating admittance y_r is correctly matched to G_0 .

As l_1 is varied the normalised admittance given by:-

$$\begin{aligned} y_2 &= y_r + j B_1 \\ &= G_r + j(B_r + B_1) \end{aligned}$$

traces a line parallel to the susceptance axis through $(G_r, 0)$, Fig. 192(b). Since l_2 is constant the corresponding admittance (y_3) traces out a circle (see (i) above).

This y_3 - circle touches

the circle $u = 0$ (the susceptance axis) corresponding to the short-circuiting of the line when $l_1 = 0$, ($n = 0.25$), the point of contact being given by $u = 0$, $n = n_2 + 0.25 = \frac{l_2}{\lambda} + 0.25$.

It also touches the circle $u = u_1$ when $B_r + B_1 = 0$

(i.e. $y_2 = G_r + j \cdot 0$),

the point of contact being

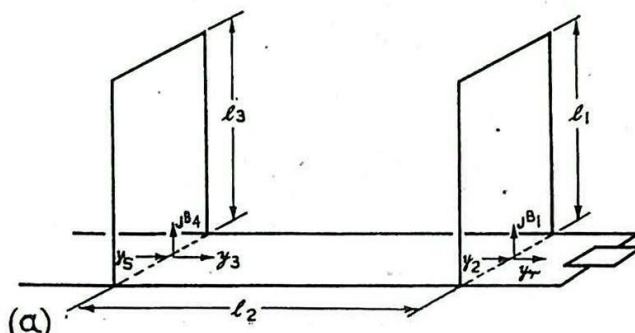
given by $u = u_1$, $n = n_2$

when $G_r < 1$ and $n = n_2 + 0.25$ when $G_r > 1$.

Actually the y_3 - circle can be shown to have centre

$$\frac{1}{2G_r} \operatorname{cosec}^2 (2\pi n_2) - j \cot (2\pi n_2)$$

and radius



LIMITS WITHIN WHICH y_s MUST LIE FOR A MATCH TO BE POSSIBLE

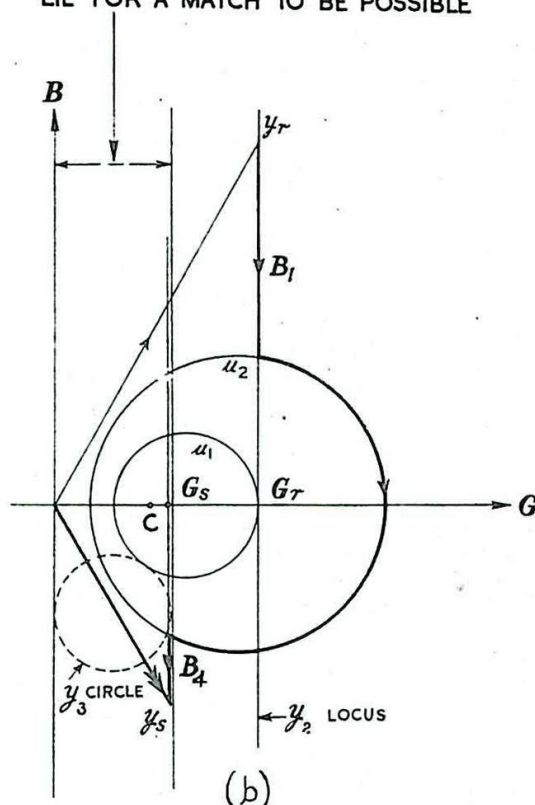


Fig. 192 - Matching by double shunt stub.

$$\frac{1}{2\overline{G_r}} \operatorname{cosec}^2 (2\pi n_2).$$

As $\overline{Y_s} = \overline{Y_3} + j \overline{B_4} = \overline{G_s} + j (\overline{B_3} + \overline{B_4})$ the possible range of $\overline{Y_s}$ available for a given $\overline{Y_r}$ is given by the vertical belt shown enclosing the $\overline{Y_3}$ -circle.

For the required value of $\overline{Y_s}$,

$$\overline{Y_3} = \overline{Y_s} - j \overline{B_4} = \overline{G_s} + j (\overline{B_s} - \overline{B_4})$$

and the locus of satisfactory values for $\overline{Y_3}$ is another line parallel to the susceptance axis, this time through $\overline{G_s}$. From an intersection, following in a counter-clockwise direction the circle $u = u_2$, which passes through the chosen point, for a distance corresponding to $n_2 = \frac{\ell_2}{\lambda}$, the appropriate value of $\overline{Y_2}$ is reached. Only one of the two possible solutions is illustrated in Fig. 192(b).

The stub lengths ℓ_1 and ℓ_3 may be determined from the normalised susceptances $\overline{B_1}$ and $\overline{B_4}$ by the method indicated in (v) above.

(vii) Matching by slugs

In general two degrees of freedom are needed in a device for matching a line to a given load, so that a single slug is not adequate. A slug may be combined with another matching device, such as a shunt stub. The following demonstration will be restricted to the particular case in which two identical movable $\lambda/4$ slugs are employed.

Suppose that two $\lambda/4$ slugs are inserted in a line of characteristic impedance R_0 so that the line with the slug present has a characteristic impedance $\frac{R_0}{m}$ ($m > 1$), as shown in Fig. 193(a).

Suppose that the standing wave ratio on the length ℓ_1 is S and that it is required to adjust ℓ_1 and ℓ_2 , if possible, so that the line at A is properly matched.

The circle diagram (b) illustrates the procedure. The circle $u = u_D$ represents the input impedance of the line to the right of D. This circle has as the extremities of a diameter the points $(\frac{1}{S}, 0)$ and $(S, 0)$. For brevity we shall denote this and similar circles as the circle $(\frac{1}{S}, S)$. The impedance $\overline{z_D}$ corresponds to a point on the circle $u = u_D$, the position of this point depending on the exact position of $\overline{z_r}$ and the value of $n_1 = \frac{\ell_1}{\lambda}$. It is necessary to convert the locus $u = u_D$ to the locus of the impedance $\overline{m z_D}$, which is the impedance of z_D

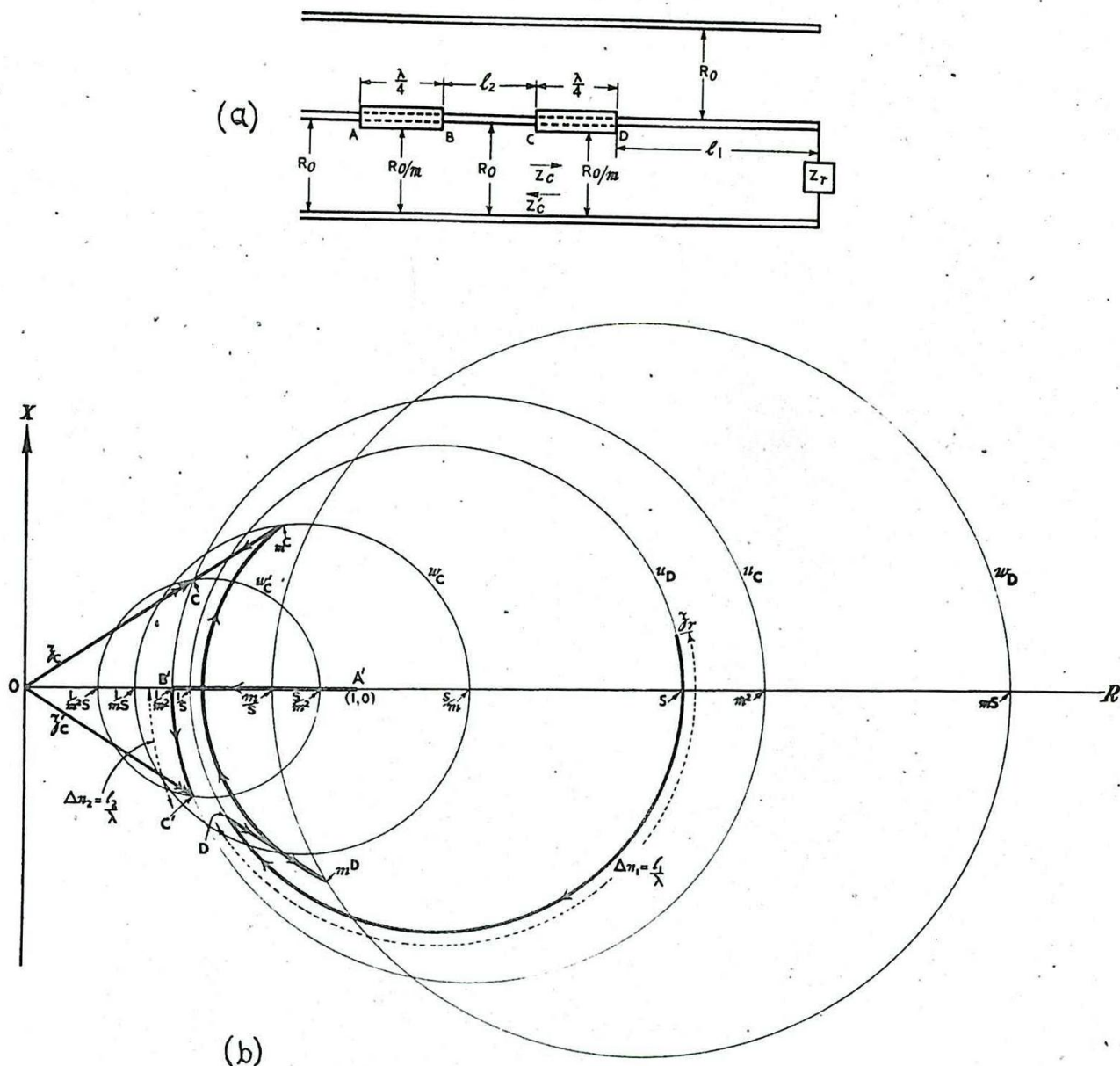


Fig. 193 - Matching by double $\lambda/4$ sliding slugs.

normalised with respect to $\frac{R_0}{m}$. This is done by multiplying each value of $\boxed{z_D}$ by m ; the circle labelled ω_D is thereby obtained (not a u-circle). The ends of its \boxed{R} -axis diameter are $(\frac{m}{S}, mS)$.

As $\boxed{mz_D}$ traverses the circle ω_D the input impedance $\boxed{mz_C} = \frac{1}{\boxed{mz_D}}$ (normalised with respect to $\frac{R_0}{m}$) at C , $\frac{\lambda}{4}$ from D , traces another circle ω_C . (See (i) at the beginning of this section). This circle is given by $(\frac{1}{mS}, \frac{S}{m})$. To convert this impedance-variation so that it represents the termination of the portion BC (normalised with respect to R_0) it is necessary to divide each impedance ω_C by m ; the locus of z_C is thereby obtained, this circle ω_C' being given by the points $(\frac{1}{m^2S}, \frac{S}{m^2})$.

Hence as the line length ℓ_1 is varied, with a constant termination $\boxed{z_r}$ the input impedance seen looking to the right at C traces the circle ω_c' .

We now consider the conditions at the sending end. The impedance to the left of A is R_0 , denoted by A' (1, 0). (The primed letters are here used to denote that the impedance is viewed from the right in Fig. 193(a), not from the left as for the unprimed letters). The impedance to the left of B is given by R_B where

$$R_0 R_B = \left(\frac{R_0}{m}\right)^2; \text{ i.e. } R_B = \frac{R_0}{m^2}, \text{ and its normalised}$$

value is $\frac{1}{m^2}$. This gives the point B' . The circle $u = u_c$ which corresponds to the standing wave pattern on the line BC is then the circle $(\frac{1}{m^2}, m^2)$, and the distance ℓ_2 from B to C (Fig. 193(a)) is given by the change in the 'n'-value of the n-arcs from B' to C' traversed in an anticlockwise direction, i.e.,

$$\Delta n_2 = \frac{\ell_2}{\lambda} = n(B') - n(C').$$

The condition for a correct match at C is that the impedances denoted by C and C' should be conjugate. C must therefore lie at the opposite end from C' of a vertical chord of the circle $u = u_c$. Hence it is necessary that the circles $u = u_c$ and $\omega = \omega_c'$ intersect. The point B' always lies inside the ω_c' -circle, since the ends of diameter of this circle are $\frac{1}{m^2 S}$, $\frac{S}{m^2}$ and B' is the point $\boxed{R} = \frac{1}{m^2}$,

S being greater than 1. Hence for intersection to occur it is necessary that the other end of this diameter, $\boxed{R} = m^2$, should lie outside the ω_c' -circle, i.e.,

$$m^2 > \frac{S}{m^2}, \text{ or } S < m^4.$$

Hence matching is possible by this method provided the standing wave ratio introduced by the mis-match at the termination is not greater than m^4 .

The various impedance transformations which occur in this method of matching are illustrated by the heavy lines in Fig. 193(b). Two solutions are possible, but only one is indicated. For simplicity the transformation due to the $\frac{\lambda}{4}$ transformer AB is

shown as a straight line from A' to B' . As described in (iii) above the impedance actually follows a u -circle (after first being normalised with respect to $\frac{R}{m}$). The distance ℓ_2 may be determined directly, as indicated, and the distance ℓ_1 can be similarly determined when the position of the load impedance $\boxed{z_r}$ is located on the circle $u = u_D$.

The problem of slug-matching is more complicated if the slugs are not $\frac{\lambda}{4}$ in length, but the procedure is the same. The loci are still circles, but the mutual relations are more complex and are not considered further.

(viii) Wide-band stub support (metallic insulator)

The arrangement is illustrated in Fig. 194(a). Suppose the characteristic resistances of the lines PQ, QR are each $\frac{R_0}{m}$, where R_0 is the characteristic resistance of the main line and $m > 1$. The circle $u = u_1$ (Fig. 194(b)) indicates the input admittance of the line QP terminated at P in R_0 for various lengths. If this line is $\lambda/4$ in length, Q is the current antinode on this u -circle. If the frequency of operation increases so that the line is slightly more than $\lambda/4$, Q appears as shown, so that the admittance \overrightarrow{OQ} contains a small inductive susceptance. If the input admittance jB of the stub QT is of the right magnitude, the resultant admittance, $\overrightarrow{OQ} + \overrightarrow{QQ'} = \overrightarrow{OQ'}$ brings the admittance back on to the circle $u = u_1$, and the input admittance of the line RQ at R, terminated as it is at Q, is the same as that at P, namely G_0 , as shown by its normalised value $\frac{1}{m}$.

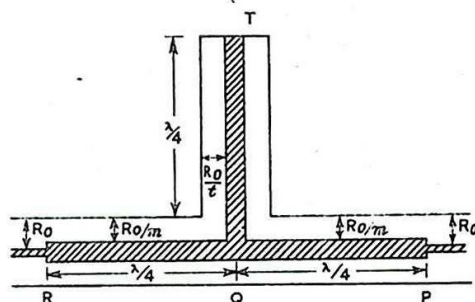
Thus, provided the change of frequency which introduces the inductive susceptance in the input admittance at Q produces just the right capacitive susceptance in the input admittance of the stub QT, a wide-band match is obtained.

If the characteristic resistance of the stub is R_0 , it may be shown that the condition for a correct match is given by

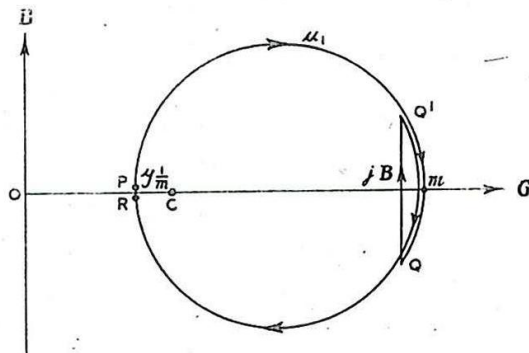
$$t = 2m(m^2 - 1).$$

A particular case arises when $t = m = \sqrt{\frac{3}{2}}$. In

this case the same thickness of inner cable may be used for both the stub and the thickened line PR.



(a)



(b)

Fig. 194 - Wide-band stub support.

CHAPTER 5

WAVEGUIDES

INTRODUCTION

1. General

We have seen in Chap. 4 that electromagnetic waves may be propagated between a pair of parallel conducting cylinders or between a pair of coaxial cylinders and that the transmission line system so formed provides a convenient means for conveying high frequency power from a source to a load. It is also possible to pass an electromagnetic wave down the inside of a hollow metal tube and, for various reasons which will be elaborated below, this is a desirable procedure at centimetre wavelengths. In place of the metal tube a dielectric rod is sometimes used. The tube or rod used for this purpose is called a waveguide. Whereas it was convenient to discuss propagation on transmission lines in terms of the voltages and currents associated with the wave disturbance it is more useful in the study of waveguides to concentrate attention on the electric and magnetic fields of the wave in the tube. It is useful, therefore, before proceeding to a detailed discussion of waveguide propagation to summarise the principle features of the electromagnetic fields of waves in free space and on transmission lines.

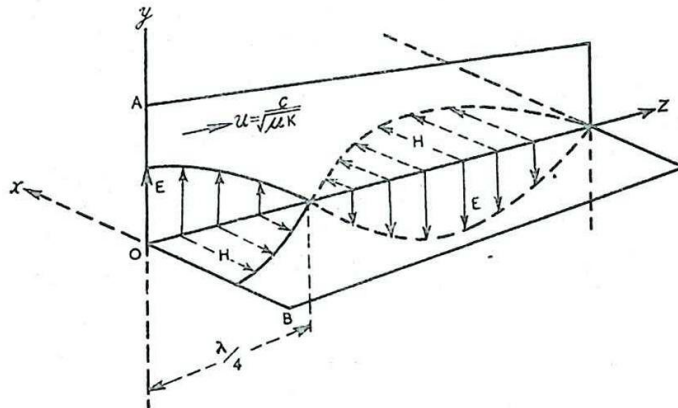


Fig. 195 - Field distributions in a plane polarised electromagnetic plane wave.

2. Properties of Electromagnetic Waves in Free Space

Fig. 195 depicts a sinusoidal plane-polarised electromagnetic plane wave travelling in an unlimited medium; its properties are as follows :-

- (i) The wave comprises oscillations of an electric field E and a magnetic field H in directions at right angles to each other and to the direction of propagation,

- (ii) the waves are transverse both in E and H and are, therefore, of the type called TEM (Transverse-Electro-Magnetic); that is, there is no component of E or H in the direction of propagation,
- (iii) The velocity of propagation in free space is equal to c, the velocity of light (3×10^8 m/sec.) In a medium of dielectric constant K and permeability μ , the velocity is

$$\frac{c}{\sqrt{K\mu}}$$

- (iv) E and H oscillate in phase with each other when the medium is non-conducting,
- (v) The amplitudes of E and H are constant all over a wave front.

3. Behaviour of an Electromagnetic Field at the Surface of A Conductor

The electromagnetic fields of waves on transmission lines and in waveguides are bounded by metal surfaces; consequently, to appreciate the forms of the field patterns it is essential to know how such fields behave at a metal surface at high frequencies.

It is assumed in the first instance that the metal has infinite electrical conductivity. It can be shown that the electric field E is perpendicular to the metal at its surface (as in electrostatics), or, in other words, the tangential component of the electric field vanishes at the surface of a perfect conductor. The magnetic field H of the wave is everywhere tangential to the surface; that is, there is no normal component of H at the surface. H is zero inside the conductor, (Fig. 196). To support the discontinuity in H at the surface, i.e., to allow the magnetic field to change suddenly from H to zero in crossing the conducting surface from without to within the conductor, there is a current sheet at the surface of the conductor.

Since any metal possesses a finite conductivity the above description of the electromagnetic field idealises the actual conditions, but the behaviour in the ideal case is a very close approximation to the actual behaviour at a metal surface when the frequency of the wave is high. At low frequencies electromagnetic fields penetrate deeply into the interior of a conducting medium and the corresponding currents are distributed throughout a volume of the medium. As the frequency is increased, however, the currents are crowded more and more into the surface of the conductor and the fields penetrate less and less deeply. This is the well known phenomenon of Skin Effect.

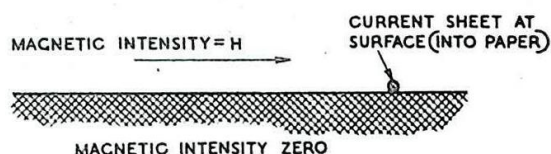


Fig. 196 - Behaviour of electromagnetic field at conducting surface.

4. Waves guided by Pairs of Conductors (Transmission Lines)

We will now discuss the properties of a wave travelling along a transmission line. They are very similar to those of the freely travelling waves in an unbounded medium already mentioned. Figs. 197 (a) and (b) represent sections of parallel pairs of long conducting cylinders whose contours are of arbitrary form but the same for each cylinder along its whole length.

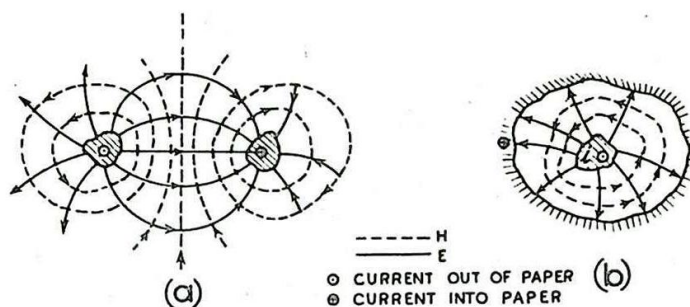


Fig. 197 - Cross-section of TEM-wave carried by conducting cylinders.

The figures indicate the patterns formed by the lines of E and H; (Since other types of wave can also be carried by the transmission lines the wave here discussed is usually called the Principal Wave). The wave system has the following properties :-

- (i) The wave is propagated parallel to the axes of the cylinders, and is transverse both in E and H; i.e., the electric and magnetic fields are each perpendicular to the direction of travel and lie in a plane parallel to the wave front as shown in Fig. 197.
- (ii) The pattern formed by the electric lines of force is that of the two-dimensional electrostatic field obtained by maintaining the cylinders (supposed infinitely long) at a suitable difference of potential. Hence, each line of force arises from a surface charge on one conductor and ends on a surface charge of opposite sign on the other conductor. The lines also cut the surface of the conductors at right angles.
(Boundary conditions, Sec. 3).
- (iii) The lines of magnetic force surround one or both conductors and form a pattern which is the same as that obtained by a suitable steady current along the cylinders but in opposite directions in each. When a high frequency wave is transmitted the current in the conductor is entirely superficial, and the contour of each conductor in the section is followed by the neighbouring magnetic lines.
- (iv) At all points the two fields are mutually perpendicular.
- (v) The velocity of the wave is the same as that for the unguided wave and is the same for all frequencies. There is no limit to the frequency of the wave which can be propagated.

5. Limitations of Cables and Advantages of Waveguides at Centimetre Wavelengths

As explained in Chap. 4 Secs. 41 - 45, one limitation to the general use of coaxial cables at centimetre wavelengths is set by excessive absorption of power. This absorption arises both from ohmic heating of the conductors and from internal loss in the dielectric, the former loss increasing in proportion to the square root of the frequency and the latter to the frequency itself. In both cases the loss is accentuated by the presence of the inner conductor. The currents responsible for the ohmic loss are equal in magnitude in the inner and outer conductors, but since they are forced by the skin effect to flow in the surface layers of these conductors they dissipate most of their energy in the inner conductor which has the smaller surface. The dielectric loss increases with the field strength, and since the field is greater near the inner conductor these losses also are concentrated there.

It is to be anticipated, therefore, that when electromagnetic waves are propagated in tubes, in which both inner conductor and dielectric are absent, the attenuation is relatively small and comparable with that due to the outer conductor of a coaxial cable of the same dimensions. Such tubes are known as Waveguides and because of the small attenuation suffered by electromagnetic waves in them they are often preferred to cables in applications using centimetre wavelengths. A dielectric rod or tube may also be used as a waveguide, but in such a case the attenuation due to the presence of the dielectric is likely to exceed that due to the finite conductivity of the metal in an air-spaced coaxial transmission line. The precise attenuation coefficient in a waveguide depends on the wavelength, cross-section and form of wave, but whereas in a good cable the attenuation coefficient for a wave of length 10 cm. is of the order of 0.5 db/metre, in a waveguide it can be of the order of 0.025 db/metre. At wavelengths of 3 centimetres the attenuation coefficient in the cable would exceed 1 db/metre and that in the guide would have risen to 0.075 db/metres. It therefore follows that cables would not be used at wavelengths of 3 centimetres except in very short lengths.

Loss of energy from a wave as it travels along a transmission line is objectionable not only because it weakens the signal which arrives at the other end, but also because of the undesirable heating produced by the wasted energy. In the case of high power transmission this heating is often sufficient to soften the polythene (softening point about 100°C) in the cable and cause breakdown.

6. Waves in Waveguides compared with those in Free Space and on Transmission Lines

The salient features of waves in guides are as follows :-

- (i) The oscillations of E and H are not both transverse and it is convenient to classify waves in guides into two principal types. Two different nomenclatures are in use for describing the waves. Sometimes the wave is known in terms of that field component which is entirely transverse; it is called, for example, a Transverse Electric, or TE wave. Sometimes the same wave is known in terms of the field component which has a longitudinal component, a wave with a longitudinal H-component, for example, being called an H-wave. The two principal types are therefore :-

- (a) H-waves
Transverse Electric
(TE) waves, }

in which the electric field E is entirely transverse but the magnetic field H possesses a component H_z in the direction of propagation (longitudinal component) as well as a transverse component H_t , Fig. 198(a). H_t

and H_z oscillate in

quadrature so that the resultant H vibration is elliptical. H_t and E vibrate in phase.

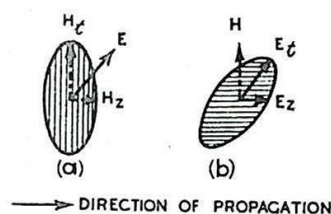


Fig. 198 - Component vectors of E-waves and H-waves.

- (b) E-waves
Transverse Magnetic (TM) waves, }

characterized by entirely transverse oscillations of H but in which E has a longitudinal component, E_z (Fig. 198 (b)). E_t and E_z oscillate in quadrature so that the resulting E vibration is elliptical.

- (ii) The amplitudes of E and H are not constant over a wave front. In the simple case of a rectangular tube they vary sinusoidally across the tube.
- (iii) The presence of the longitudinal component of either the E or the H field causes the wave front to travel faster than the energy of the wave. The velocity of the wave front is called the Phase Velocity u_p , whilst the speed with which energy is transmitted along the guide is called the Group Velocity u_g . It may be shown that in all cases :-

$$u_p \cdot u_g = c^2.$$

The wavelength λ_g of a wave in a guide exceeds the wavelength λ of the free space wave of the same frequency since :-

$$u_p = f \lambda_g \quad \text{and} \quad c = f \lambda.$$

- (iv) Electromagnetic disturbances are propagated as waves inside a metal tube only if the free space wavelength ($\lambda = c/f$) is less than a critical value λ_c , called the Cut-off Wavelength. The corresponding frequency $f_c = c/\lambda_c$ is called the Cut-off Frequency. The cut-off wavelength is determined by the type of wave and the geometry and dimensions of the cross-section of the tube, but in a given tube the greatest value of λ_c is of the order of magnitude of the larger cross-sectional dimension of the guide.

- (v) The following relation exists between the wavelength λ_g of a wave in a guide, its cut-off wavelength λ_c and the free-space wavelength λ :-

$$\frac{1}{\lambda_g^2} = \frac{1}{\lambda^2} - \frac{1}{\lambda_c^2};$$

multiplication by $1/f^2$ gives :-

$$\frac{1}{u_p^2} = \frac{1}{c^2} - \frac{1}{f^2 \lambda_c^2}.$$

The phase velocity u_p in the guide is evidently a function of the frequency; whereas the velocity in free space is the same for all frequencies.

WAVES IN A RECTANGULAR TUBE

7. Synthesis of an H-Wave

Most of the important features of waves travelling in guides can be derived quite easily from a consideration of the case of a rectangular guide carrying a simple form of H-wave known as the H_{01} wave. This derivation is so instructive that it will be given in detail and the important points brought out. For more complicated cases results only will be quoted and no detailed derivation will be given. There are no important fundamental points which cannot be seen from the simple treatment now to be given, and it should not be thought that any complicated "mathematical" analysis can give any deeper insight into the phenomena.

The method of approach is as follows. We shall show that if a train of plane electromagnetic waves is incident obliquely on a perfectly conducting sheet of metal then the incident and reflected waves combine to produce outside the metal a composite wave in which the motion appears to take place in a direction parallel to the surface of the metal. This moving wave system can, if we wish, be thought of as a peculiar type of wave guided by a single infinite metal sheet. Precisely the same wave disturbance is produced when a wave is incident at an angle on to a plane surface of the earth which can be considered to be perfectly conducting. A detailed discussion of this case of an incident wave combining with its reflected wave after reflection from the surface of the earth is given in Chap. 17. We shall also show that in this type of wave the amplitudes of the electromagnetic fields alter as we cross the wave front and moreover the fields are not entirely transverse to the direction of advancement of the wave. If now we wish to insert in the guided wave another metal sheet parallel to the first to form an opposite boundary of a waveguide we find in general that it will disturb the wave pattern. It is found, however, that a metal sheet can be inserted without disturbing the wave system in certain planes where the magnitudes and directions of the fields are suitable. Under these conditions we have a wave system which is guided between two parallel planes which can be thought of as two opposite sides of a waveguide. We next show that the insertion of a second pair of plates perpendicular to the first pair will not disturb the wave, and thus we have built up a type of wave which can travel in a rectangular guide.

We now conduct the synthesis in detail. Let AB represent a metal sheet of infinite conductivity and suppose that a train of plane waves incident upon it in air (ideally, vacuum) dielectric, Fig. 199 (a), gives rise to an equal reflected wave (b).

The diagrams illustrate the distribution of fields in space at any chosen instant. The full lines are wave fronts over which E and H have maximum values; over the dotted lines mid-way between, E and H are everywhere zero. The section through the wave fronts is so chosen that the magnetic vectors H (represented by arrows on the full lines) lie in the plane of the paper, and the electric vectors E (represented by circles with a dot or cross inside) stand perpendicular to the plane of the paper. λ is the free-space wavelength, and the arrows c indicate the directions of the incident and reflected waves. Between one full line and the next there is a complete reversal of both E and H fields, corresponding to a phase shift of 180° . Consequently, the separation of the full lines is $\frac{\lambda}{2}$.

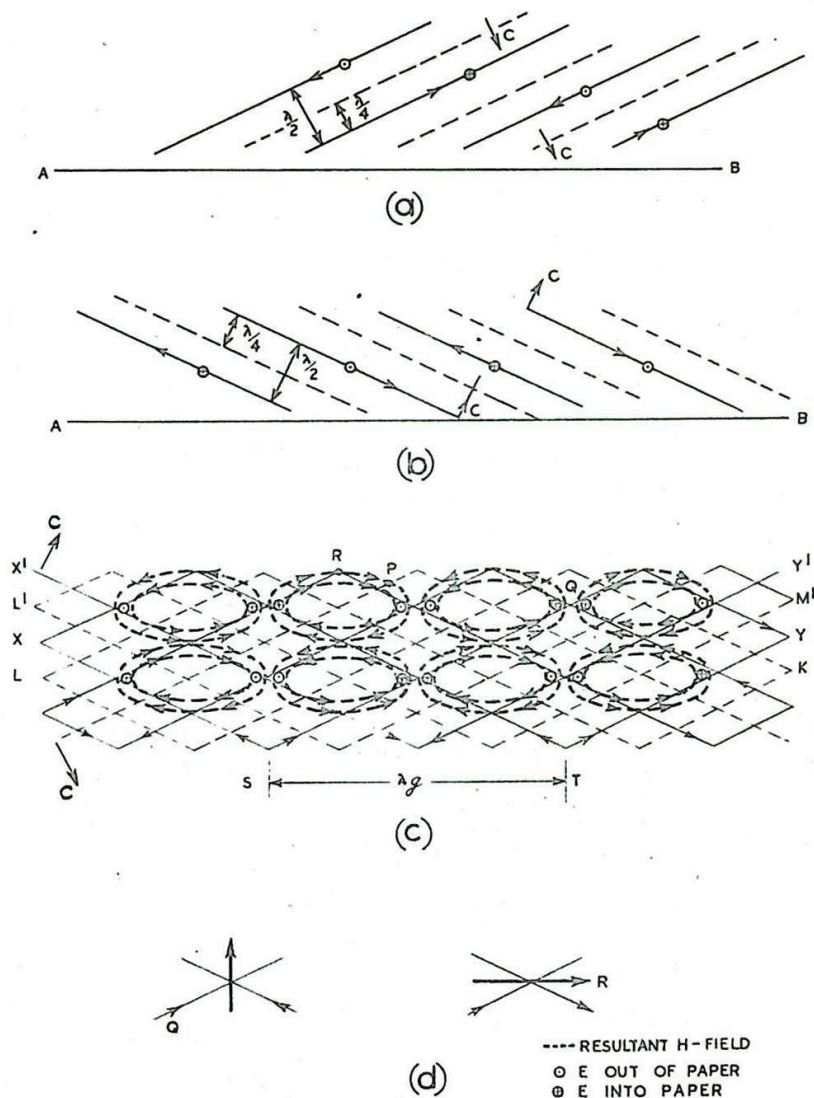


Fig. 199 - Synthesis of an H_{on} -wave.

As the time t changes the patterns move in the directions indicated by the arrows perpendicular to the wave fronts and with the velocity of light c . In Fig. 199 (c) the two wave trains are shown passing across the same region of space by superimposing the patterns shown at (a) and (b). The c -arrows are moved to the extremities of the wave fronts in order not to confuse the diagram. The resultant H-field at any point is obtained by adding the two vector magnetic fields of the

original wave trains. The direction of the resultant H-field is indicated by the arrows on the heavy lines. At certain points of the pattern the direction of the resultant field may be seen by inspection. For instance, at all points on a dotted wave front the fields of that wave train are zero and the resultant field is that of the other wave train. Thus, along a dotted line the direction of H is parallel to the wave fronts of the other wave system, as for example at the point P. Where two dotted lines cross there is no field whatsoever. At a point of intersection of two full lines the resultant field H is directed along the bisector, either of the acute, or of the obtuse angle between the full lines according to the relative directions of the arrows appropriate to the intersecting wave fronts. The points Q and R of Fig. 199 (c), shown separately in Fig. 199 (d), provide examples.

When the directions of the resultant magnetic field are indicated by arrows at a sufficient number of points it is possible to sketch the lines of magnetic force in the pattern of the resultant field. These are seen to be closed curves whose precise form is determined by the angle of incidence of the original wave. As the centre of each magnetic loop is a point of intersection of two dotted lines there is zero resultant magnetic or electric field there.

Since the electric fields of the elementary plane waves shown at (a) and (b) are everywhere directed either up or down perpendicular to the plane of the paper, the resultant electric field in the pattern shown at (c) is also normal to the plane of the paper and attains a maximum strength twice that of the field in the separate wave trains. The circles with a cross or dot within indicate the sense of the local resultant electric field. It is to be understood that the pattern extends in depth into and out of the plane of the paper.

8. Pattern Velocity (Phase Velocity) of H-waves

A fixed feature of the pattern shown in Fig. 199 (c), such as the resultant magnetic field at the point Q, is associated with a point of intersection of two wave fronts of the wave trains shown at (a) and (b). The two wave fronts intersecting at Q are shown as QF and QG in Fig. 200.

As the wave fronts QF and QG travel in the ray directions FK and GK respectively, at velocity c , their point of intersection Q travels at velocity u_p in the direction LM'; moreover Q moves from Q to K in the time taken for F to move from F to K or G from G to K. Let the angle FQK (equal to GQK) be α ; then, since the angles FQK and KQK are right angles, it follows that :-

$$\frac{u_p}{c} = \frac{QK}{FK} = 1/\cos \alpha = \frac{\lambda_g}{\lambda} \quad (\text{Sec. 6 (iii)}),$$

or

$$u_p = c/\cos \alpha = c \sec \alpha.$$

It is convenient at this point to summarise what has already been deduced in this section. We have seen that when a plane electromagnetic wave is incident obliquely on a metal sheet it combines with its reflected wave to synthesise a new type of wave in which the electric and magnetic field patterns are quite different from those of the constituent plane polarised waves. The direction of propagation of the new wave is parallel to the plane of the metal and the pattern velocity u_p is greater than the velocity c . It is shown that $u_p = c/\cos \alpha$, where α is the angle of incidence of the original wave on to the metal.

Let us now examine in more detail the nature of the electric and magnetic fields over the wave system. Over planes such as XY and $X'Y'$ (Fig. 199 (c)) which separate two layers of magnetic loops, the magnetic field is everywhere tangential to the plane and the electric field is zero at all points of the plane. Such planes are nodal planes of E . Conversely, at the intervening planes such as LM , $L'M'$ the magnetic field is perpendicular to these planes and the electric field reaches its maximum values. The distance ST , containing two complete loops as shown, is the wavelength λ_g

of the composite pattern since the pattern repeats identically in a displacement of this amount parallel to XY . The composite pattern shown in Fig. 199 (c) refers to a chosen instant of time, but as the patterns of the elementary waves shown at (a) and (b) move obliquely with a velocity c , so is the composite pattern displaced without distortion in the direction of symmetry XY at another velocity u_p .

The magnetic field comprises a system of closed loops and consequently possesses in general a component H in the direction of propagation; the electric field E is entirely transverse but its amplitude varies in a direction perpendicular to the direction of propagation XY and lying in the plane of H . This wave is of the type called an H-wave in Sec. 6.

The H-wave is, as yet, of unlimited extent and it is necessary to discover whether such a wave can be caused to travel down a metal tube. We proceed in stages by considering first the possibility of propagating an H-wave between a pair of parallel metal plates of unlimited extent.

9. Propagation Between Parallel Plates and in Rectangular Tubes

As already mentioned, over the nodal E -planes XY , $X'Y'$, etc. (Fig. 199 (c)) which separate adjacent layers of magnetic loops, the electric vector vanishes, and the magnetic vector is tangential to these planes. Reference to Sec. 3 shows that over these nodal planes, the behaviour of the electromagnetic field of the H-wave is the same as that at the surface of a perfect conductor. If, therefore, a perfectly conducting plate is inserted into the wave pattern of Fig. 199(c) to coincide with a nodal E -plane such as XY , the pattern above the plate is undisturbed and that below may be removed. The pattern above is supported by transverse surface currents i as

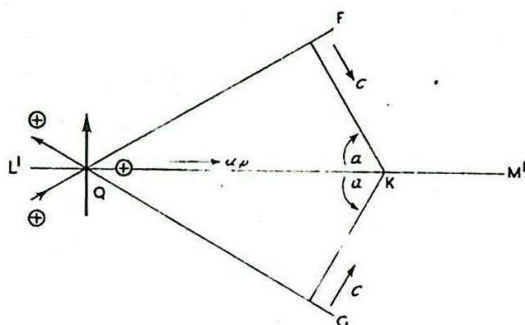


Fig. 200 - Representation of pattern (phase).

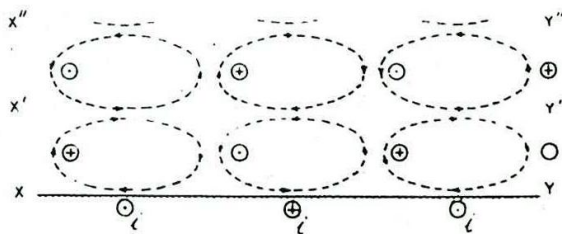


Fig. 201 - Positions of nodal planes.

indicated in Fig. 201.

It is, therefore, possible to insert a second metal sheet at any of these planes to form with the first a pair of sides of a waveguide. The second plate may be inserted so that it coincides with one of the more distant nodal planes such as $X'Y'$ or $X''Y''$. In this way any desired number of layers of loops of Fig. 199 (c) and 201, may be trapped between the plates to give an H-wave propagated at velocity u_p tangential to the plates but confined to the region between them. The elementary waves of Fig. 199 (a) and 199 (b) may now be considered to undergo successive reflections at the plates. Fig. 202 depicts a three-layer H-wave propagated between a pair of parallel plates at a separation b but of unlimited lateral extent.

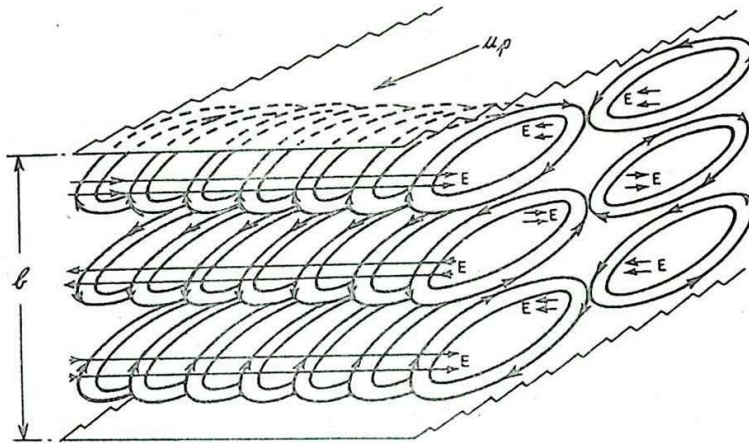


Fig. 202 - Three-layer H-wave.

To convert the parallel plate system into a rectangular tube it is necessary to insert side walls in such a way that the boundary conditions to be satisfied at the wall surface are consistent with the preservation of the wave pattern. The walls must evidently be inserted as shown in Fig. 202, that is, the wall surfaces are parallel to the planes containing the magnetic loops. Consequently H is everywhere tangential to these walls and, as indicated in the figure, E is perpendicular to them. The separation a in the present instance can be given any value.

An H-wave of the nature described above is not the only type of wave that can be propagated in a rectangular tube and it is therefore convenient at this stage to give the H-wave under discussion its usual designation. A wave of the type shown in Fig. 203, but with a number n of complete layers of magnetic loops in the ' b ' dimension is called an H_{0n} (or TE_{0n}) wave. That shown in Fig. 203 is, therefore, an H_{03} wave.

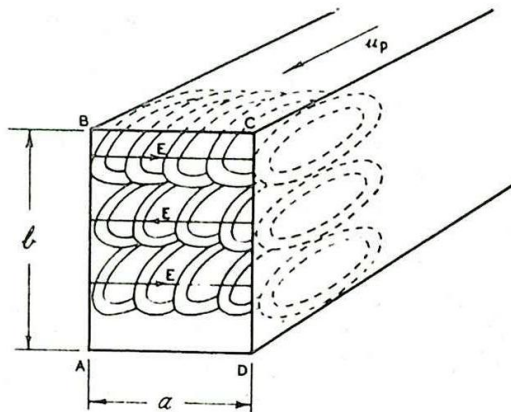


Fig. 203 - H_{03} -wave in rectangular guide.

The interpretation of the suffix zero is as follows :- if we choose any starting point on the face AB (Fig. 203) and move parallel to AD across to the opposite point on the wall CD, no variation in any component of H or E occurs. The suffix zero indicates this independence of E and H of displacements parallel to the edge 'a' of the tube section. On the other hand, in moving parallel to AB from the face AD to the face BC n layers of H-loops are crossed (or, more accurately, n half-cycles of variation in any one of the components of E or H are observed).

Of the possible H_{0n} -waves in rectangular waveguides one is of outstanding practical importance; it is the H_{01} -wave and it will be further discussed below.

10. The H_{01} -wave in a Rectangular Guide

This wave, as the suffixes indicate, contains only a single layer of H-loops between the faces AD and BC (Fig. 203) and corresponds to the case in which conducting plates are inserted at adjacent nodal planes XY and X'Y' (Fig. 201). Fig. 204 represents a section parallel to both the H-field and the direction of propagation of an H_{01} -wave in a rectangular waveguide.

The wave fronts (e.g. AC) of the elementary plane waves reflected from plate to plate are shown, as well as a magnetic loop of the resultant H_{01} -wave.

The ray direction corresponding to the wave front AC is BO and it is supposed that this ray makes an angle of elevation with the reflecting planes, as shown. The distance BO is $\lambda/4$ where λ is the free-space wavelength of the waves, and $CO = \lambda_g/4$, where λ_g is the wavelength of the H_{01} -wave in the guide.

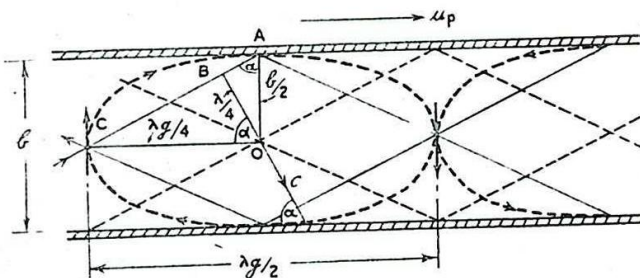


Fig. 204 = Section of H_{01} -wave in a rectangular guide.

From the diagram we deduce :-

$$\sin \alpha = BO/AO = \lambda/2b \quad \dots \dots \dots (1),$$

$$\cos \alpha = BO/CO = \lambda/\lambda_g = c/u_p \quad \dots \dots \dots (2).$$

Squaring both (1) and (2), and adding, gives

$$(\lambda/2b)^2 + (\lambda/\lambda_g)^2 = 1,$$

so that

$$\left(\frac{1}{\lambda_g}\right)^2 = \frac{1}{\lambda^2} - \frac{1}{(2b)^2} \quad \dots \dots \dots (3).$$

Equations (1), (2) and (3) are easily adapted to cover the case of an H_{0n} -wave by replacing b everywhere by b/n , since the distance b now spans n layers of magnetic loops instead of a single layer, and OA becomes $b/2n$.

Thus, $\sin \alpha = n \lambda / 2b \dots\dots\dots (1a),$

$\cos \alpha = \lambda / \lambda_g = c / u_p \dots\dots\dots (2a),$

$1 / \lambda_g^2 = 1 / \lambda^2 - (n / 2b)^2 \dots\dots\dots (3a).$

11. Cut-Off Wavelength of the H_{01} -wave

If a wave which has a wavelength λ in free space is to be fitted into a tube by reflection back and forth between faces we have seen that it must be inclined so that the distance between nodal planes is equal to $1/n$ th of the distance b between the faces where n is any integer.

Fig. 205 shows that if $\frac{1}{2} \lambda$ is less than b this can be done for $n = 1$. As λ decreases, the steepness of incidence of the wave will decrease and all wavelengths down to zero can be fitted into the tube. As λ tends to zero the original wave tends to travel along the surface of the metal. Incident and reflected waves tend to coincide and the pattern velocity approximates to that in free space. As the free-space wavelength increases the original wave has to be more steeply incident on the surface of the metal. Reference to Fig. 200 shows that the pattern velocity increases and finally, when the original wave is incident normally on the metal, the pattern velocity is infinite. Under these conditions the incident and reflected waves combine to give a standing wave system with nodal planes at the two metal surfaces. This limit is reached when the distance b is equal to $\frac{1}{2} \lambda$, and waves which have wavelengths greater than

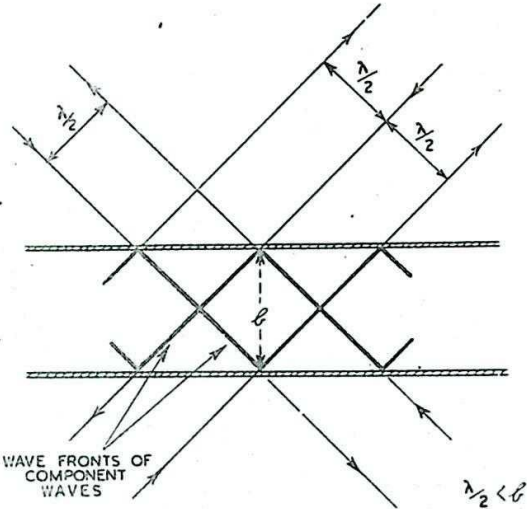


Fig. 205 - H_{01} -wave: relative dimensions.

$\lambda_c = 2b \dots\dots\dots (4)$

cannot be propagated down the guide. The corresponding frequency, given by

$f_c = \frac{c}{\lambda_c} = \frac{c}{2b} \dots\dots\dots (5)$

is called the Cut-Off Frequency for the guide and the corresponding free space wavelength λ_c is called the Cut-Off Wavelength. The cut-off wavelength for an H_{0n} -wave is given by

$\lambda_c = \frac{2b}{n} \dots\dots\dots (4a).$

We have already deduced the following relation (3a) between the free-space wavelength and the pattern wavelength in the guide for an H_{on} -wave :-

$$\left(\frac{1}{\lambda_g}\right)^2 = \left(\frac{1}{\lambda}\right)^2 - \left(\frac{n}{2b}\right)^2,$$

and we now see that this can be written :-

$$\left(\frac{1}{\lambda_g}\right)^2 = \frac{1}{\lambda^2} - \frac{1}{\lambda_c^2} \dots\dots\dots (6),$$

which is an expression of very general validity.

12. Evanescent Modes

It is of interest to enquire what kind of disturbance results if we put electromagnetic energy into a guide with a frequency corresponding to a free space wavelength of greater than λ_c . We have just seen that no wave-motion is possible. What then does happen?

In order to examine this problem let us write the equation for a wave of frequency f and wavelength λ_g in the complex exponential form where y is the magnitude of the disturbance at a distance x from the origin.

$$y = y_0 \ e^{2\pi j\left(ft - \frac{x}{\lambda_g}\right)} \dots\dots\dots (7).$$

In the case where the free space wavelength is greater than the cut-off wavelength ($\lambda > \lambda_c$) the equation (6) shows us that $\frac{1}{\lambda_g^2}$ is negative so that $\frac{1}{\lambda_g}$ is a purely imaginary quantity which may be written:-

$$\frac{1}{\lambda_g} = \pm j\alpha.$$

The wave motion of equation (7) then becomes :-

$$\begin{aligned} y &= y_0 \ e^{2\pi j\left(ft \pm j\alpha x\right)} \\ &= y_0 \ e^{\pm 2\pi\alpha x} \cdot e^{j2\pi ft} \dots\dots\dots (8) \end{aligned}$$

This expression (8) represents not a travelling wave motion but a disturbance in which the fields at all distances (x) oscillate in the same phase with the frequency f (factor $e^{j2\pi ft}$) but their amplitudes die away exponentially with increasing distances (factor $e^{\pm 2\pi\alpha x}$).

A disturbance of this type is called an Evanescent Mode. Evanescent modes are important for the understanding of the behaviour of obstacles and irregularities in waveguides. The exponential diminution of amplitude with distance also finds an important practical application in the piston attenuator (Sec. 24).

13. Cut-Off Wavelength of Different Modes of H-Waves

Expression (4a) shows that the higher the order of mode (i.e. the larger n) the smaller becomes its cut-off wavelength λ_c in

the tube of given dimension b . When λ exceeds λ_c for a certain mode in a given guide any electromagnetic disturbance in this mode must be of the evanescent type.

In a given tube there are no modes with cut-off wavelengths greater than that of the H_{01} -mode. The mode is therefore of great practical importance, for it is possible to convey power down a guide of suitable dimensions by the H_{01} -wave alone and to exclude other modes except where they may be excited in an evanescent form by obstacles or at the source. It is thus possible to radiate from the end of a guide into a mirror in a predictable manner because the distribution of field across the end of the guide is due to the H_{01} -wave alone. There are also other advantages, connected with switching and matching, in retaining only a single travelling wave in the guide.

As a practical example, we consider a waveguide which has been used in practice at a wavelength of about 9 cm. Its sectional dimensions are $a = 1''$; $b = 2\frac{1}{2}''$, (Fig. 203). Thus, corresponding to dimension a , the cut-off wavelength of the H_{01} -mode is $a\lambda_c = 2'' = 5.1$ cm., and to dimension b , $b\lambda_c = 5'' = 12.7$ cm. It is possible, therefore, to propagate in this guide an H_{01} -wave at $\lambda = 9$ cm. only if the plane of its magnetic loops lies parallel to the edge b , since $a\lambda_c < \lambda < b\lambda_c$. The polarisation of the E vector in the wave is not, therefore, ambiguous. With $\lambda = 9.1$ cm. and $2b = 12.7$ cm. we find $\alpha = 45^\circ$ and $\lambda_g = 13$ cm. The corresponding cut-off wavelengths of the H_{02} mode are $a\lambda_c = 1'' = 2.54$ cm. and $b\lambda_c = 2\frac{1}{2}'' = 6.35$ cm. This and higher modes are, therefore, evanescent for waves of $\lambda = 9$ cm.

14. Method of Launching an H_{01} -wave in a Rectangular Waveguide

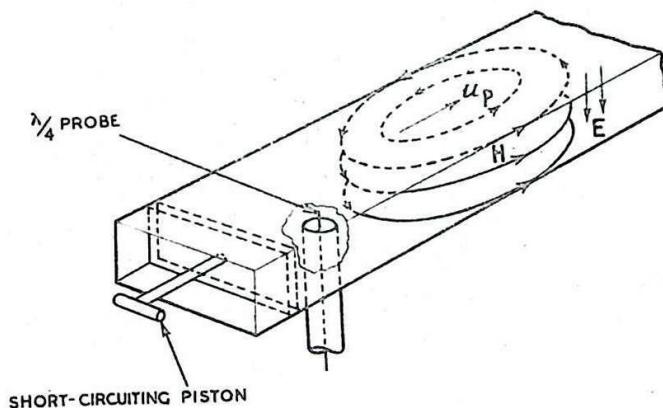


Fig. 206 = Launching an H_{01} -wave.

Fig. 206 indicates the commonest method of launching an H_{01} -wave in a rectangular guide. A probe, usually the extension of the inner conductor of a short length of coaxial transmission line, protrudes through the centre of one of the broad walls of the guide a distance $\lambda/4$ (not $\lambda_g/4$) into the guide. It then radiates as an aerial into the guide space. As it is not generally required to divide the stream of power, the guide to one side of the probe is closed by a short-circuiting piston whose position can be adjusted to cause the short length of guide between it and the probe to act as a matching reactance. The correct position of the piston will be roughly $\lambda_g/4$ from the probe. Near the probe the electromagnetic

field is complicated and consists of an H_{01} -mode with evanescent modes superimposed, but at a sufficient distance down the guide the field simplifies to that of a travelling H_{01} -wave, as indicated in Fig. 206.

15. Wall Currents in the case of the H_{01} -wave

Surface currents flow on the walls of the guide to support the tangential components of the magnetic field of the wave pattern at the metal surface. We consider the case of the H_{01} -wave.

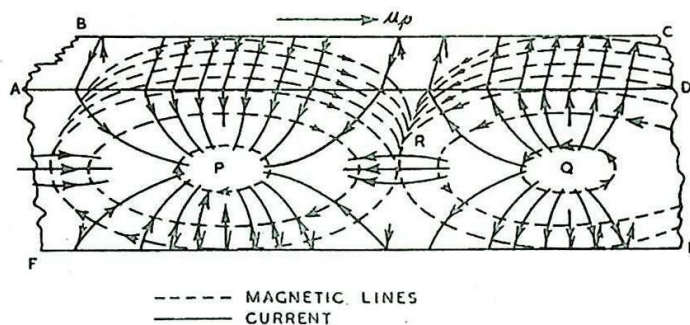


Fig. 207 - Wall currents: H_{01} -wave.

Fig. 207 illustrates the instantaneous directions of flow of the surface currents in the walls of a rectangular guide in relation to the magnetic field. The current flows everywhere at right angles to the magnetic field at the walls. Thus, on the face ADEF to which the magnetic loops are parallel, the currents converge on the area around P and diverge from Q. Thus positive charge is beginning to accumulate around P and negative charge around Q. There are no fields, however, at P and Q. The region around R carries negative charge and the opposite region on the opposite face of the guide carries positive charge. The transverse electric field of the wave is a maximum at the section R but is zero at P and Q. The currents on the wall ABCD and the opposite wall are entirely transverse, as shown, and the whole current pattern is carried along with the rest of the wave pattern at speed u_p .

It is important to appreciate the pattern of current flow in the walls when deciding where a hole or slot may be placed. In a standing-wave indicator, for instance, it may be necessary to cut a long slot in one of the walls parallel to the axis of the guide in order to insert a probe into the field. If the slot were placed in the wall ABCD or its opposite, and ran parallel to AD, it would clearly run across the currents in the wall. The flow of the currents would be disturbed and the slot would radiate into space. The slot may however be cut in the centre of the face ADEF along the line FRQ, so that there are no currents perpendicular to it, and if the slot is narrow the wave within the tube is unaffected by the slot. The radiation of energy from slots in waveguides is further considered in Chap. 17. Secs. 54 - 56.

16. E_{nm} -waves and H_{nm} -waves

By following the procedure of Sec. 7 we may also synthesise an E_{0n} -wave capable of propagation between a pair of parallel conducting

planes from a pair of simple plane polarised electromagnetic waves. It will suffice to indicate the procedure.

We begin with a plane wave incident obliquely on a metal as in Fig. 199 but with the E and H positions of the lines interchanged. Thus, in Figs. 199 (a) and (b) we suppose the vectors E to lie in the plane of the paper and H to stand perpendicular to the plane of the paper. Consequently, in the modified diagram the arrows along the wavefronts represent the direction of E and the circles with a cross or dot within represent H. With the directions of c as shown, however, when the arrows are also drawn as shown, it is necessary to reverse the sense of the H vectors from those indicated by the circles.

In the composite diagram Fig. 199 (c) the closed loops now represent the E-lines and the circles (with their senses reversed) the H-lines. The wave, is, therefore, an E-wave (TM wave). To imprison n loops of this wave pattern between a pair of parallel conducting plates it is again necessary to insert the plates into the pattern in positions such that the electromagnetic boundary conditions are not infringed. Since the closed loops are now E-lines and the transverse lines those of H, the correct positions for the plates are any two of the planes LM, L'M', etc., over which the tangential component of E vanishes and H is entirely transverse. A cross-section of an E_{on} -wave between two plates is shown in Fig. 208 (a), and a view into the advancing wave in Fig. 208 (b).

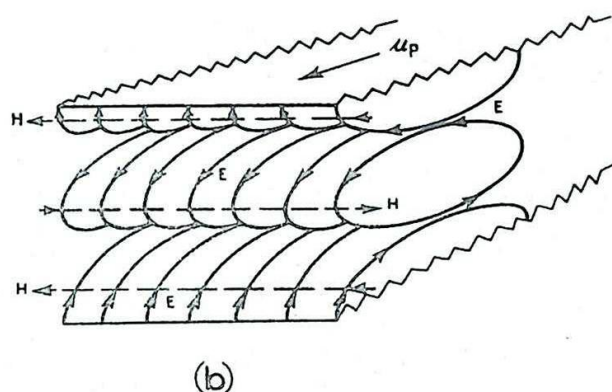
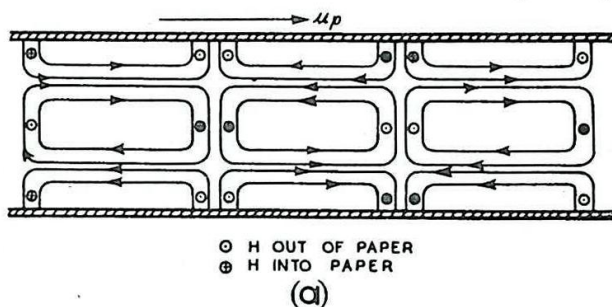


Fig. 208 - E_{on} -wave between parallel plates.

The formulae for cut-off wavelength λ_c and wavelength λ_g between the plates are the same as those for the H_{on} -wave, i.e., equations (1a) to (4a) in sections 10.

When, however, an attempt is made to transform the pair of plates into a tube by putting in side walls the wave system is disturbed since H is normal to these walls and E entirely tangential. Thus, there is no E_{on} -wave in a rectangular tube. The simplest E-wave in a rectangular tube is the E_{11} -wave whose pattern is shown in Fig. 209, but is not derived. Fig. 209 (a) represents a central section of the guide containing the axis and parallel to one of the walls. The electric lines of force are in the form of loops which meet the walls perpendicularly. Along the axis the lines are grouped into bundles. Fig. 209 (b) is a transverse section of the wave pattern at the position P in (a) with the wave approaching the observer. The

centre of the section is a source of electric lines which diverge from an axial bundle and terminate on the walls.

The magnetic field is entirely transverse and consists of closed loops surrounding the central bundle of electric lines as shown. The pattern in the section at Q (a) is the same as that shown at (b) but with the directions of the fields everywhere reversed. The central region of the section, therefore, is a "sink" of electric lines instead of a source.

One method (but not the simplest) of deducing the form of the E_{11} -wave pattern is to build up the wave from an original plane wave sent into the tube in a suitable direction so that successive reflections occur from all four walls. The plane of polarisation of the original wave must also be chosen appropriately.

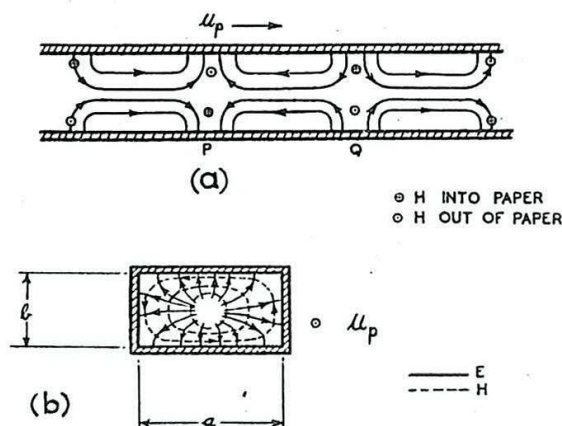


Fig. 209 - E_{11} -wave in a rectangular tube.

It can be shown that if a and b are the cross-sectional dimensions of the guide then the wavelength λ_g is given by :-

$$\frac{1}{\lambda_g^2} = \frac{1}{\lambda^2} - \left\{ \left(\frac{1}{2a} \right)^2 + \left(\frac{1}{2b} \right)^2 \right\} \dots \dots \dots (9).$$

The cut-off wavelength may be deduced from equation (1), since when $\lambda = \lambda_c$ the wavelength in the guide is infinite.

Hence

$$0 = \frac{1}{\lambda_c^2} - \left\{ \left(\frac{1}{2a} \right)^2 + \left(\frac{1}{2b} \right)^2 \right\},$$

i.e.

$$\frac{1}{\lambda_c^2} = \left(\frac{1}{2a} \right)^2 + \left(\frac{1}{2b} \right)^2 \dots \dots \dots (10).$$

The general E_{nm} -wave (TM_{nm}) may be considered as a number nm of E_{11} -waves squeezed into a single tube so that individual patterns fit together to form a single large pattern. For instance, Fig. 210 shows a transverse section of the field pattern of an E_{32} -wave at a position where the E-lines are diverging from the axial bundles. This takes the form of an overall pattern in which the unit is the E_{11} distribution. Further, in any pair of adjacent unit patterns the one has a source at its centre P and the other a sink Q. By interposing metallic partitions these patterns could be isolated into nm independent E_{11} -waves. The suffix n is usually associated with the

horizontal dimension a , and m with vertical dimension b . It may be inferred from formula (2) above that, since the equivalent sectional dimensions of one of the constituent E_{11} -waves are a/n and b/m , the cut-off wavelength of the E_{nm} -wave is given by :-

$$\frac{1}{\lambda_c^2} = \left(\frac{n}{2a}\right)^2 + \left(\frac{m}{2b}\right)^2 \dots\dots\dots (11).$$

Like the E_{nm} -wave, the H_{nm} -wave can also be considered as an assemblage of unit patterns of a simpler wave, in this case the H_{11} -wave whose sectional pattern is shown in Fig. 211. The corners of the section are sources and sinks of H-lines which run down the tube as bundles located in the corners. Each magnetic line forms a closed loop. The electric lines are entirely transverse as shown.

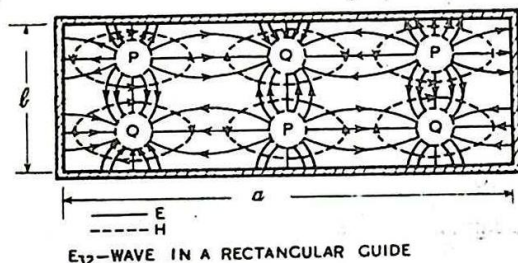


Fig. 210 - E_{32} -wave in a rectangular guide

WAVE ADVANCING OUT OF PAPER

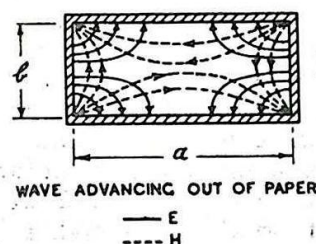


Fig. 211 - H_{11} -wave in a rectangular guide.

The transverse pattern of the H_{nm} -wave comprises nm units each like Fig. 211, in which the fields in adjacent units are reversed. The resultant pattern is not shown.

The formula for the cut-off wavelength of an H_{nm} -wave is the same as that for an E_{nm} -wave, viz., equation (11).

Of the possible E_{nm} - and H_{nm} -waves in rectangular guides only the H_{01} has found extensive practical application, for reasons already given in Sec. 13. Consequently the properties of the general E_{nm} - or H_{nm} -waves will not be further discussed.

WAVES IN CIRCULAR TUBES

17. General

Waves can be propagated in guides whose cross section is other than rectangular. The most commonly used of these other types has a circular cross section. A more detailed treatment of the possible modes of propagation in a circular guide is complicated, but a good insight into the main phenomena involved is obtained by imagining the circular guide to be built up from a deformed rectangular guide of the type previously considered.

18. H_{01} -wave in a Circular Guide

We begin with an H_{01} -wave propagated between a pair of parallel conducting plates as shown in Fig. 212 (a). The plates are next

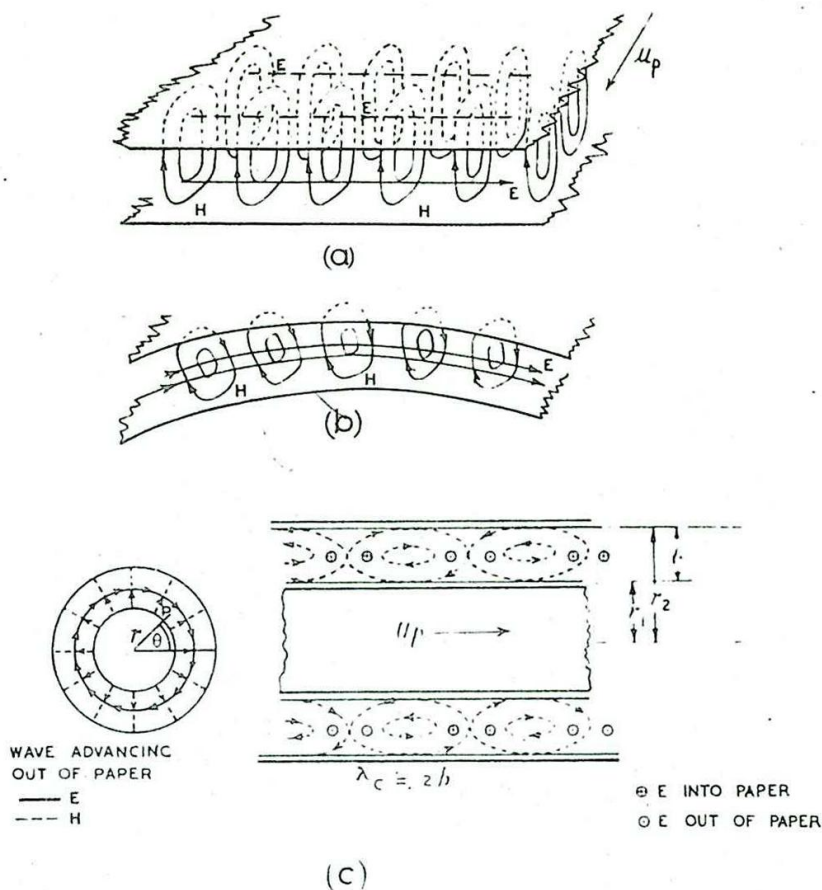


Fig. 212 - H_{01} -wave on a coaxial line.

curved to form a coaxial transmission line, the one plate forming the inner cylinder and the other the outer as shown at (b) and (c). The H_{01} -wave pattern between the plates transforms, on curvature, to an H_{01} -wave between the coaxial cylinders as depicted at (c). This figure represents a type of wave which can be propagated along a coaxial line used as a waveguide. It should not be confused with the entirely different pattern of the TEM wave (Sec. 2), or the principal mode of propagation in a coaxial line.

Suppose the inner cylinder of the coaxial system of Fig. 212 (c) were to shrink, so that it becomes an axial wire. The magnetic loops now virtually touch on the axis but are in fact separated by the wire. Since there is no discontinuity in H at the surface of the wire, no current is needed to support the H -field so that the wire is superfluous and may be removed. There remains a hollow tube with a wave travelling along it. This is the H_{01} -wave for a circular waveguide (Fig. 213). We can obtain a rough idea of the magnitude of its cut-off wavelength as follows. The cut-off wavelength of the wave in the original rectangular guide is $2b$ and as a result of the deformation of the rectangle the dimension b becomes approximately equivalent to the radius r_g of the circular guide.

Hence we expect the cut-off wavelength to be approximately $2r_g$. Exact calculation shows it to be $1.64 r_g$.

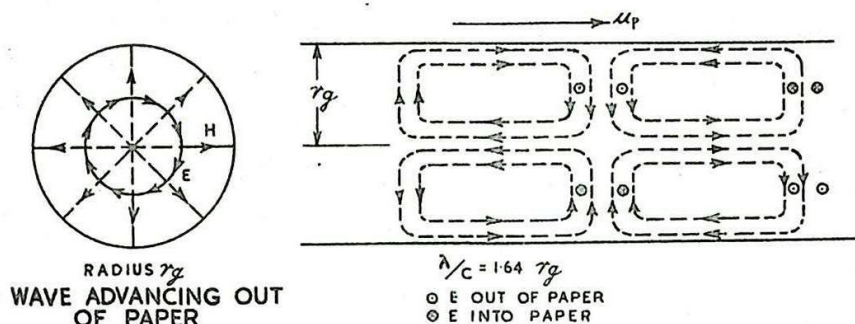


Fig. 213 - H_{01} -wave in a circular guide.

19. Classification of Waves in Circular Guides

It is convenient at this point to remark on the significance of the suffixes in the appellation H_{01} of the wave which has just been described, and to generalise the system of nomenclature to deal with H_{nm} - and E_{nm} -waves.

In rectangular guides the suffix zero indicates that none of the field components was changed in a displacement parallel to one of the edge (edge a) of the section (horizontal displacement in Fig. 211). In general, each suffix indicates the number of repetitions of the unit pattern (apart from reversal of field directions) included in the appropriate sectional dimension.

In the case of waves in circular tubes, cylindrical co-ordinates are required. Thus, the position of a point P, Fig. 212(c), within a section of the wave pattern is given in terms of its polar co-ordinates (r, θ). A horizontal displacement across the pattern between the plates of Fig. 212(a) transforms to a circular displacement of P on a circle of constant radius r . Thus, the suffix zero indicates no variation in any component due to a change in θ at a fixed r .

An H_{-} -wave between coaxials is similarly obtained by transforming an H_{0n} -wave between plates (Fig. 202) into a cylindrical pattern as described above. This becomes an H_{0n} -wave in a circular guide when the radius of the inner cylinder is reduced to zero. The pattern of this wave also exhibits axial symmetry, but contains n layers of H-lines in the radial direction.

We note that the suffix zero of an H_{0n} - (or E_{0n}) wave between cylinders or in a circular tube denotes that the wave possesses axial symmetry.

As for waves in a rectangular guide, the symbol H indicates that there is a longitudinal component of H in the wave propagated along the guide, and similarly the symbol E indicates a longitudinal component of E.

An E_{nm} -wave, or an H_{nm} -wave has $2n$ units of the basic pattern in the variation of θ from 0 to 2π (i.e. it is divided by n

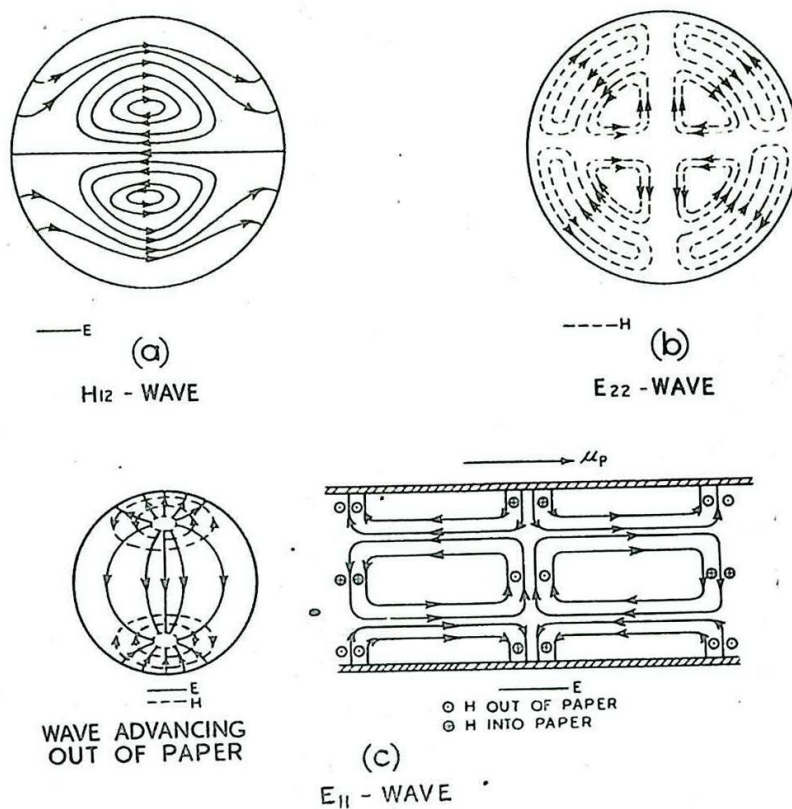


Fig. 214 - Waves in circular guides.

diameters) and m units in the variation of r from 0 to r_g , where r_g is the radius of the guide. As examples of the system of classification Fig. 214 shows cross-sectional drawings of an H_{12} -wave (a) and an E_{22} -wave (b) in a circular guide. Fig. 214 (c) shows transverse and longitudinal sections of an E_{11} -wave.

20. The E_{01} -wave in a Circular Guide

This wave is the analogue of the E_{11} -wave in a rectangular guide shown in Fig. 209. If the guide be given a square section which is then distorted into a circular section, the pattern of Fig. 209 transforms into that of Fig. 215.

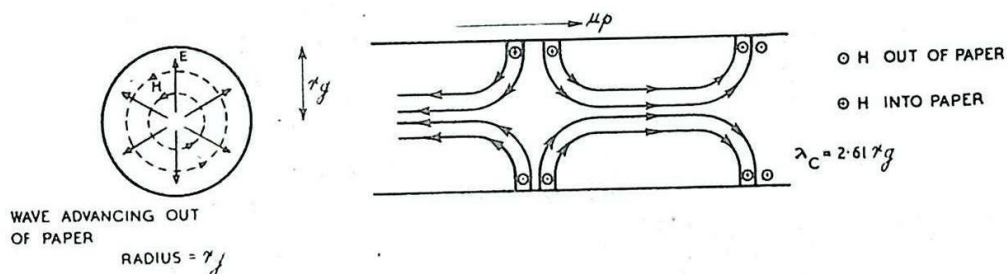


Fig. 215 - E_{01} -wave in a circular guide.

The electric lines are loops whose longitudinal portions form axial bundles which diverge radially to the walls as shown. The magnetic field is entirely transverse and in the form of circular lines about the source of divergence of the E-lines.

The cut-off wavelength of the E_{11} -wave in a square section tube of side b is $\lambda_c = b\sqrt{2}$. If we make the crude identification of radius r_g of the circular section with $b/2$, half the side of the square, the following approximate formula for the cut-off in the circular guide is obtained :-

$$\lambda_c = 2r_g\sqrt{2} \doteq 2.8 r_g .$$

The correct formula is $\lambda_c = 2.61 r_g$.

A method of launching the E_{01} -wave in a circular guide is indicated in Fig. 216. The wave is radiated from a probe through the centre of a terminating plate at the end of the guide. The E_{11} -wave in a rectangular guide may also be launched in this fashion.

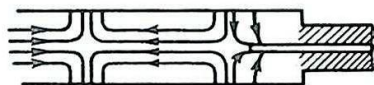


Fig. 216 - Launching an E_{01} -wave in a circular guide.

The E_{01} -wave is of considerable practical importance since it is commonly employed in rotating joints for use with scanners at wavelengths of 3 cms. Its importance lies in the fact that of the waves with axial symmetry in a circular guide it possesses the greatest cut-off wavelength. Further, it is easily excited by an H_{01} -wave in a rectangular guide feeding through a junction in the wall of the circular guide. It is, however, more convenient to discuss the details of rotatable joints in a separate section (Sec. 42).

21. The Septate Guide

We consider an H_{01} -wave in a rectangular guide polarised as shown in Fig. 217 (a).

The guide is again curved into a coaxial system but this time it is the edge b which is curved. The two original side walls are retained and merged into a single metallic septum which supports the inner cylinder. This structure is called a Septate Waveguide. The field of the wave is distorted as indicated in the figure. The original flat magnetic loops are bent over so that their longitudinal portions run in opposite senses one on either side of the septum. There is thus a discontinuity in H in crossing the septum which therefore carries surface currents to support this discontinuity. The electric field in the wave is entirely radial. It vanishes on each face of the septum and is a maximum diametrically opposite the septum.

The feature of the septate guide that recommends it for practical employment is its relatively large cut-off wavelength. The cut-off wavelength of the prototype guide of Fig. 217 (a) is $\lambda_c = 2b$. The dimension b transforms into the mean circumference of the septate whose cut-off wavelength is therefore given by :-

$$\lambda_c \doteq 2 \pi (r_1 + r_2).$$

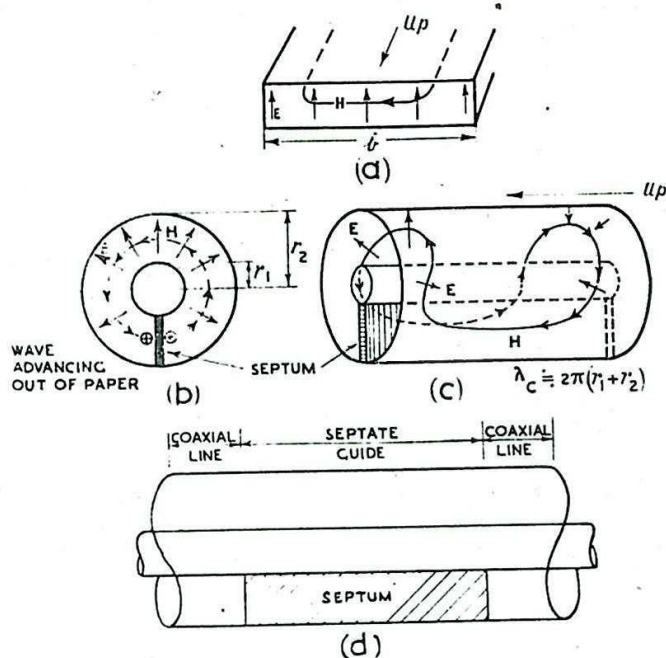


Fig. 217 - Septate guide.

This cut-off wavelength is approximately twice as great as the longest cut-off wavelength of any wave in a circular guide of the same outside radius $r_g = r_2$.

The similarity of the sectional pattern (Fig. 217(b)) of the septate wave to the pattern in the principal wave of a coaxial line permits the septate wave to be easily fed from a coaxial transmission line formed as an extension of the inner cylinder without the septum. The septate pattern also feeds efficiently into a coaxial line (d). Since the principal wave in a coaxial line possesses axial symmetry, the combination of a septate guide with short lengths of coaxial line at each end may be used to provide a rotatable joint at wavelengths of the order of 10 cms because the E_{01} -wave requires a waveguide whose radius is inconveniently large at these wavelengths.

The combination of septate guide and coaxial line is virtually a coaxial line in which the dielectric support for the inner is replaced by the septum. The combination is capable of handling higher powers than a polythene cable.

22. The H_{11} -Wave in a Circular Guide

The stages in the development of the H_{11} -mode, by the method of the preceding sections, are shown in Fig. 218. A pair of rectangular guides carrying identical H_{01} -waves (a) are bent until they merge into a coaxial line with two septa, (c). The progressive distortion of the field may also be followed from the diagrams. Since the magnetic lines on opposite sides of each septum are in the same direction and of equal density, the septa are superfluous and may be removed without affecting the field. The field pattern of

this coaxial mode is that shown at (c).

The value of λ_c for each of the prototype guides of (a) is $\lambda_c = 2b$:

The analogue of $2b$ in (c) is the mean circumference. Thus, for this coaxial mode we have

$$\lambda_c \doteq \pi(r_1 + r_2). \quad \text{The}$$

electric field is entirely transverse and oppositely directed at diametrically opposite points of the section. Of supplementary modes in coaxial cables this mode possesses the greatest cut-off wavelength and is consequently the one that is most readily excited.

To obtain an H_{11} -wave in a circular guide the inner cylinder of the coaxial is allowed

to shrink to zero. The lines of electric force join to bridge the section and the magnetic lines move inwards towards the diameter to give the pattern of figure (1). The inner of the coaxial may be dispensed with since there is no discontinuity in the E-field at the centre and no currents are required there to support the wave. The magnetic lines are loops whose longitudinal portions are collected into bundles above and below but which spread across the section where the magnetic field is transverse.

According to the formula for λ_c in the coaxial line we should expect to obtain λ_c in the guide by putting $r_1 = 0$.

$$\text{Thus, } \lambda_c \doteq \pi r_2.$$

The correct formula is $\lambda_c = 3.42 r_g$ where r_g is the radius of the guide.

Fig. 219 illustrates one method of launching an H_{11} -wave in a guide. The inner conductor of the coaxial line is carried across the guide and terminates in another coaxial line with an adjustable tuning piston. This piston together with the movable terminating piston in the guide itself serves to match the guide to the line. Another method employs a simple $\lambda/4$ probe as indicated in Fig. 206.

The H_{11} -wave is the

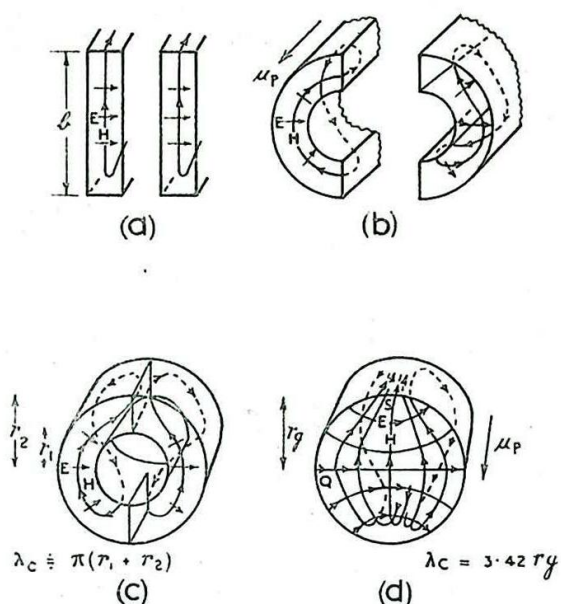


Fig. 218 - H_{11} -wave in a circular guide.

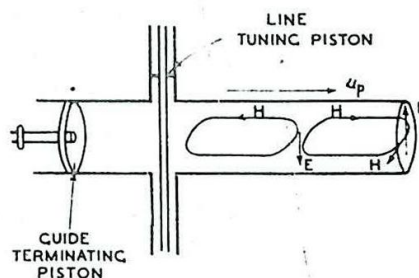


Fig. 219 - Launching an H_{11} -wave.

analogue of the H_{01} -wave in a rectangular guide. It is often used because it possesses the largest value of λ_c of any mode in a circular guide.

ATTENUATION IN WAVEGUIDES

23. Attenuation of Propagated Modes

It has been supposed, in what has preceded, that the walls of the waveguide were perfectly conducting, and consequently that progressive waves were propagated in it with no loss of power. In practice, however, the metal walls possess a large, but finite, electrical conductivity and the currents that flow in them when an electromagnetic wave is propagated along the guide are accompanied by the generation of heat. The energy transformed into heat is abstracted from the power carried by the wave whose field components are therefore attenuated with axial distance z away from the source. In the expressions for the amplitude of each component of the electric field E and of the magnetic field H there must now be included a factor $e^{-\alpha_l z}$. The ratio of the field amplitudes at two corresponding positions on the cross-section at distance z and $(z + d)$ is therefore $e^{-\alpha_l d}$ and the logarithm of this ratio to the base e , is $\alpha_l d$; this is the loss, measured in nepers, suffered by the wave. The loss per unit length is therefore α_l nepers or $\alpha_l = 8.686 \alpha_l$ decibels. The loss coefficient depends on the electrical conductivity of the material of the walls, the frequency, the mode of wave being transmitted and the dimensions and geometry of the waveguide. When the loss ($\alpha_l \lambda_g$ nepers) suffered in a distance equal to the wavelength λ_g is very much less than unity, as occurs in practice in waveguides with silver, copper or brass walls, the appropriate formulae for α_l may be derived by the following method.

The field pattern is assumed to be the same as for a waveguide formed of perfectly conducting material, with the exception that the amplitudes of all field components are exponentially attenuated in passage down the guide. In other words, apart from this longitudinal attenuation, it is assumed that in waveguides with highly conducting metal walls, the field components in the wave are the same as those in a hypothetical waveguide with perfectly conducting walls; whereas in fact there is a small component of the electric intensity E tangential to the metal surface.

The components of E and H each contain the factor $e^{-\alpha_l z}$, consequently the power W transmitted over a cross section of the waveguide, which is obtained by integrating the product of the transverse components of E and H over the cross section, contains the factor $e^{-2\alpha_l z}$

Thus,

$$W = W_0 e^{-2\alpha_l z}$$

where W_0 is the power transmitted and

$$\frac{dW}{dz} = -2\alpha_l W.$$

But the loss of power per unit length, $-\frac{dW}{dz}$, is the energy dissipated in the walls of the waveguide in ohmic heating. This energy loss per

unit length can be calculated separately from the currents in the walls which are directly proportional to the tangential components of the magnetic field at the wall, and in terms of the conductivity of the metal and the frequency. Calling this energy loss per unit length A , we have

$$A = - \frac{dW}{dz} = 2\alpha_l W.$$

or,

$$\alpha_l = \frac{A}{2W} \quad \text{nepers per unit length.}$$

Thus, the loss coefficient is found theoretically by calculating the energy lost per unit length in ohmic heating, and the flux of power W over the corresponding cross section. This procedure leads to the following formulae for attenuation coefficients of progressive waves in rectangular waveguides.

The symbols which appear in these formulae have the following interpretations :-

f is the frequency of the wave in cycles per second.

f_c is the cut-off frequency for the wave mode concerned.

a and b are the dimensions of the waveguide cross section measured in metres; the dimension a is associated with the mode integer n and b with m .

R_s is the surface resistance in ohms per unit square of the metal surface; that is, power is dissipated in unit area of the surface at the rate $\frac{1}{2} R_s I^2$ watts, where I is the RMS value of the total surface current in amperes per unit length. R_s is found from the following formula:

$$R_s = 2 \pi (10^{-7} \rho \cdot f)^{\frac{1}{2}} \quad \text{ohms per unit square....(1)}$$

where ρ is the specific resistance of the wall metal in ohms per metre cube.

For copper, $\rho = 1.6 \times 10^{-8}$ ohms per metre cube; consequently, with copper walls,

$$R_s = 8 \pi \times 10^{-8} \sqrt{f}.$$

The surface resistance of any other metal may be obtained by multiplying this value by the factor

$$N = \sqrt{\frac{\rho_{\text{metal}}}{\rho_{\text{copper}}}}.$$

α_l is the loss coefficient in nepers per metre. To obtain the loss in decibels per metre it is necessary to multiply the value in nepers by 8.686.

Formulae for loss coefficients in rectangular air-filled Wave Guides :

The H_{10} -Wave

(Electric Field parallel to the edge b of the cross section)

$$\alpha_f = \frac{R_s}{120 \pi b} \cdot \frac{1}{\sqrt{1 - (f_c/f)^2}} \cdot \left[1 + 2 \frac{b}{a} \left\{ \frac{f_c}{f} \right\}^2 \right]$$

nepers per metre (2).

To obtain α_f for the H_{01} -mode, interchange a and b in this formula.

The E_{nm} -Wave

$$\alpha_f = \frac{2R_s}{120 \pi b} \cdot \frac{1}{\sqrt{1 - (f_c/f)^2}} \cdot \left[\frac{m^2 (b/a)^3 + n^2}{m^2 (b/a)^2 + n^2} \right]$$

nepers per metre (3).

The H_{nm} -Wave

$$\alpha_f = \frac{2R_s}{120 \pi b} \cdot \left[\frac{\frac{b}{a} (m^2 \cdot \frac{b}{a} + n^2)}{(m^2 (\frac{b}{a})^2 + n^2)} \cdot \left\{ 1 - (f_c/f)^2 \right\}^{\frac{1}{2}} + (\rho + 1) \left\{ \frac{f_c}{f} \right\}^2 \right] \times \frac{1}{\sqrt{1 - (f_c/f)^2}}$$

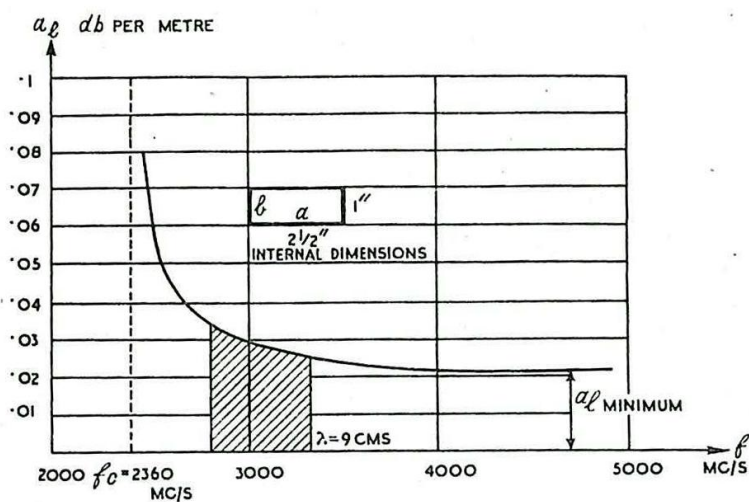
nepers per metre (4).

According to these formulae the attenuation in a rectangular guide depends on

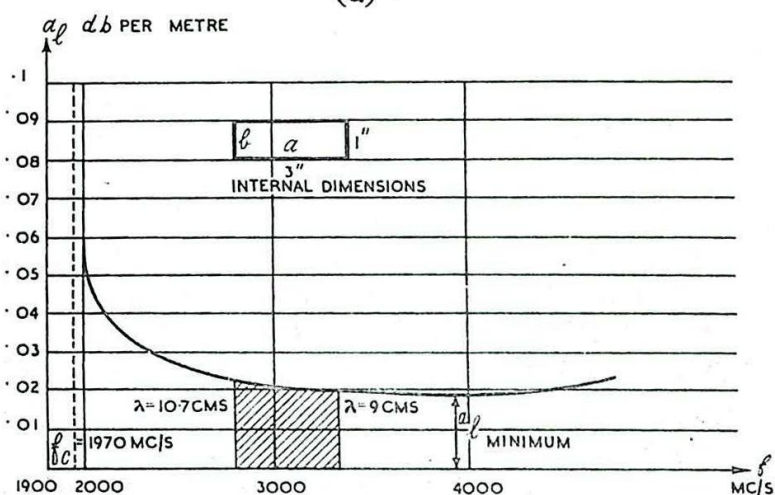
- (a) the size of the tube,
- (b) the ratio of the cross-sectional dimensions,
- (c) the resistivity of the wall ,
- (d) the frequency.

We shall not, however, discuss the influence of these factors in detail, because in radar practice the only wave that is employed to carry power through a long run of waveguide is the H_{01} -wave in a rectangular guide and it will suffice to discuss the attenuation of this wave in the actual $2\frac{1}{2}'' \times 1''$ and $3'' \times 1''$ waveguides used in service equipments.

The losses in copper waveguides with these standard internal dimensions are calculated as a function of frequency from formula (2). The results are exhibited in the curves of Figs. 26 (a) and (b) which give the loss α_f in decibels per metre suffered by an H_{10} -wave, as a function of frequency f .



(a)



(b)

Fig. 220 - Loss coefficient a_l in standard S-band copper waveguides as a function of frequency.

According to formula (2) the loss becomes infinite at the frequency $f = f_c$ due to the factor $\frac{1}{\sqrt{1 - (f_c/f)^2}}$

and also at $f = \infty$ due to the factor \sqrt{f} . We may therefore expect a to reach a minimum value, in a given waveguide at some frequency greater than f_c . These properties of the curves are shown Figs. 220 (a) and (b). The asymptotic approach of a_l to infinity as f approaches f_c and the minimum for each curve are indicated.

The values for a_l relate to waveguides with copper walls, but to obtain the loss when another metal is used it is necessary only to multiply the ordinates of the curve by a factor $N = \sqrt{\frac{\rho_{\text{metal}}}{\rho_{\text{copper}}}}$,

where ρ is the specific resistance.

Thus :-

METAL	BRASS	SILVER	ALUMINIUM
N	2.1	0.97	1.27

With iron, the factor N is

$$N = \sqrt{\frac{\rho_{\text{iron}}}{\rho_{\text{copper}}}} \times \text{permeability of iron}$$

Because of the high permeability of iron and steel the attenuation in tubes of these metals is large.

The position of the band of frequencies corresponding to wavelengths from 9 - 10.7 cm. is indicated on each curve.

The smaller $2\frac{1}{2}'' \times 1''$ waveguide, to which Fig. 220 (a) relates, is used at a wave-length of about 9 centimetres, that is at the right hand end of the band shown in the figure, where the attenuation is about 0.027 N db, per metre. For equipments working at wave-lengths at the other end of the band, from 10 - 11 cms, this waveguide would be operated on the rising portion of the curve (a). Consequently it is desirable to employ a guide of larger dimensions for which the 10 - 11 cm band is nearer the loss minimum. Curve (b) shows how this is achieved by the use of a 3" by 1" guide. The attenuation at the operating frequencies in both waveguides is therefore approximately 0.025 N db. per metre and in a brass waveguide this is $0.025 \times 2.1 = 0.0525$ db. per metre.

The corresponding loss in the standard polythene cable, Uniradio 21, is about 0.6 db per metre; that is, about ten times as great.

Since the attenuation is proportional to the surface resistance it is important to avoid corrosion of the interior surface of the guide due to weather or salt spray. For this reason waveguides are often plated with silver or cadmium internally, or hermetically sealed at their free ends by cellophane diaphragms, the air within being kept dry with silica-gel cells which communicate with the interior through small holes in the walls.

24. Attenuation of Evanescent Modes

Consider an arbitrary disturbance, sinusoidal in time, to be excited in a tube whose dimensions in cross-section are small enough to make the longest cut-off wavelength of possible modes much less than the free-space wavelength of the disturbance. The arbitrary disturbance may, in principle, be represented as a set of co-existent evanescent modes whose amplitudes and phases are appropriately chosen. Each mode (nm) decays in amplitude with distance z according to the factor

$$e^{-\delta_{nm}z} \text{ where } \delta_{nm} = 2\pi \left[\frac{1}{(\lambda_{c,nm})^2} - \frac{1}{\lambda^2} \right]^{\frac{1}{2}}$$

and is the loss coefficient of the mode of order nm. We have postulated however that λ greatly exceeds $\lambda_{c,nm}$; consequently,

$$\delta_{nm} \doteq \frac{2\pi}{\lambda_{c,nm}}$$

Thus, the smaller the cut-off wavelength $\lambda_{c, \text{min}}$ the more rapidly does the mode attenuate with distance. At a sufficient distance z from the source of the disturbance only the mode with the largest cut-off wavelength persists, provided its relative amplitude is appreciable at the source. This means in practice that the modes of higher order disappear first.

We conclude, therefore, that at a sufficient distance z , depending on the precise circumstances, the amplitude of all the field components of the residual disturbance decay with distance according to a simple exponential law, i.e.

$$e^{-\delta z},$$

where $\delta = 2\pi/\lambda_c$, λ_c being the cut-off wavelength of the most persistent mode. Further, the coefficient δ is independent of the free-space wavelength λ . If the magnitude of the EMF induced by the field in a loop feeding into an output coaxial cable is V_1 at $z = z_1$, but V_2 at $z = z_2$, with the loop similarly situated with respect to the field in the two cases, then

$$\frac{V_2}{V_1} = e^{-\delta(z_2 - z_1)}.$$

The reduction of signal strength in distance $(z_2 - z_1)$ is:

$$\log_e \left(\frac{V_1}{V_2} \right) = \delta(z_2 - z_1) \text{ nepers},$$

or:

$$20 \log_{10} \frac{V_1}{V_2} = 8.686 \delta(z_2 - z_1) \text{ db}.$$

Thus, reduction in signal strength measured in decibels (or nepers), depends only on the displacement $(z_2 - z_1)$ of the loop in a given tube, δ being a constant, $2\pi/\lambda_c$, independent of frequency. This result is in effect correct for all frequencies which are low compared with the cut-off frequency.

We have, therefore, a method of reducing a signal, say that at $z = z_1$, by any desired number of decibels by a simple linear displacement of a pick-up loop.

This method therefore provides a simple means of producing a voltage which is a given fraction of the input voltage, and this fraction can be altered by a linear displacement of the pick-up loop. Moreover, for wavelengths about 10 cms or less such an attenuator is compact and is suitable for inclusion in test equipment. A device of this nature is called a Piston Attenuator.

CAVITY RESONATORS

25. General

It is found that electromagnetic oscillations can be excited within an empty cavity bounded by conducting walls in much the same way as hollow gas-filled vessels can be excited into acoustical resonance. A given cavity resonator will resonate at a number of discrete frequencies each corresponding to a particular mode of oscillation with its own characteristic electromagnetic field pattern. These field patterns, like those of progressive waves in waveguides, can be classified into E and H types.

Since the free-space wavelength that corresponds to the mode with the lowest frequency is of the order of magnitude of the greatest linear dimension of the cavity, it follows that at centimetre wavelengths a cavity resonator has a small physical size which renders it convenient for laboratory use and for incorporation in equipments. The principal uses for cavity resonators are :

- (i) tuned elements in oscillators in place of the conventional L-C-R circuits which are physically unrealisable at centimetre wavelengths.
- (ii) accurate wave metres.
- (iii) echo boxes, which are equivalent to ringing circuits.

The electro-magnetic oscillations in cavity resonators in the form of hollow rectangular boxes or cylinders have field patterns, with one exception, that are the same as those belonging to standing waves in rectangular and circular waveguides and it is therefore possible to employ elementary methods to deduce many of the important features of cavity resonators.

We begin with the simple case of a rectangular resonator.

26. Stationary Waves on a Waveguide

Consider a rectangular waveguide in which a complete standing wave is produced by allowing two trains of H_{01} -waves with the same wavelength λ_g and field strengths, to be propagated in it in opposite directions. The progress of the individual waves is indicated successively in the first two of each of the Figs. 221 (a), 221 (b) and 221 (c). The third diagram in each case gives the resultant field derived by superimposing the field patterns of the two waves. We take the time $t = 0$ to correspond to the instant at which the magnetic loops in the two oppositely travelling patterns superimpose exactly, so that at $t = 0$ in the standing wave the magnetic field strength is a maximum at all points as indicated in the third of Figs. 221 (a).

Because the constituent waves travel in opposite senses, the electric fields are in opposition when the magnetic fields are additive. Consequently, the electric fields cancel at $t = 0$, and at this instant the field in the standing wave is entirely magnetic. Fig. 221 (b) illustrates the situation at time $t = T/4$, one quarter of a cycle later, T being the period of the oscillations. The travelling wave patterns in Fig. 221 (b) are displaced one quarter wavelength,

$\lambda_g/4$, to the right and left respectively, as can be seen from the displacement of the pair of loops which has been distinguished by a horizontal arrow at the centre. When the patterns are superimposed the magnetic fields cancel but the electric fields add to produce a maximum. Thus at $t = T/4$ the field in the standing wave is entirely electric. Similarly, from Fig. 221 (c) we deduce that at $t = T/2$ the electric field vanishes and that the field is again entirely magnetic but reversed in direction, compared with the field at $t = 0$. At $t = 3T/4$ we should again find the resultant field to be entirely

electric but reversed in direction relative to that at $t = T/4$. At time $t = T$ the field is again that of Fig. 221 (a). At other instants in the cycle the electric and the magnetic components are both present. Further, whatever the field intensity at any instant, the directions of the E and H fields remain fixed and the pattern does not progress to right or left as in the constituent progressive H_{01} -waves. We therefore note the following features of the electromagnetic field for a

standing wave.

- (i) The field patterns of the magnetic and electric fields are individually the same as those in the progressive waves.
- (ii) Whereas in the progressive wave the field patterns are propagated along the axis of the waveguide at speed u_p , but the field intensities in the moving patterns are unchanged, in the stationary wave the patterns are fixed in position but the field intensities oscillate harmonically between maximum positive and negative values.

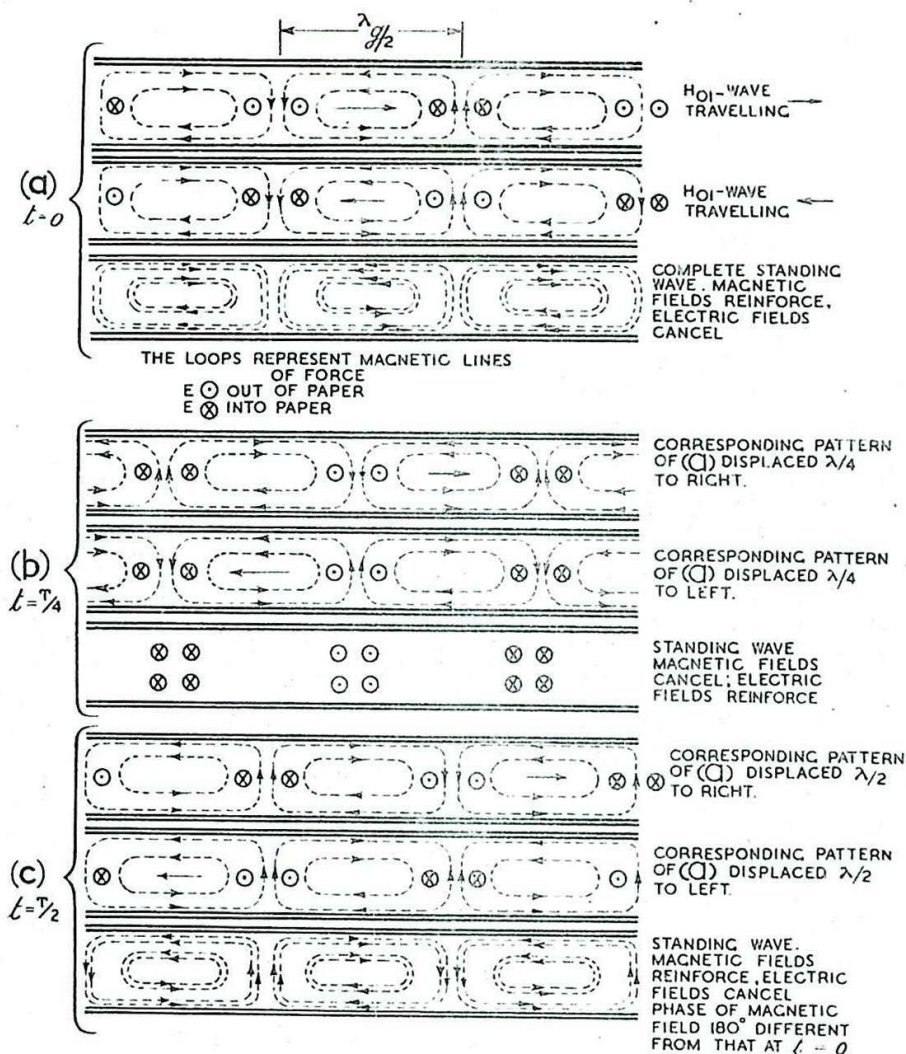


Fig. 221 - Superposition of two oppositely travelling H₀₁-waves of equal strength to produce a complete standing wave.

- (iii) In the progressive waves the transverse components of the electric and magnetic fields coincide (near the ends of the largest loops), but in the stationary wave the positions at which the transverse components of the magnetic and electric fields have maximum amplitudes are separated by a quarter of a wave-length ($\lambda_g/4$).

(The electric field is concentrated around the centre of a magnetic loop).

- (iv) The resultant electric and magnetic fields within each $\lambda_g/2$ cell of the pattern oscillate in quadrature.

There is no mean flow of power along the axis of the waveguide, but the energy stored in each cell of the pattern is transformed every quarter-period from the magnetic to the electric form, and back again in the succeeding quarter-period. The oscillations of the electro-magnetic field are therefore analogous to the mechanical vibrations of a pendulum or an escapement wheel in which the energy is transformed alternately from one to the other of the kinetic and potential forms.

27. Field Patterns in Cavity Resonators

We have obtained a stationary but oscillating electro-magnetic field; we now consider whether this field is one that can exist within a closed cavity with conducting walls. At the walls the field must satisfy the following conditions: the magnetic field cannot intersect any portion of the conducting surface of the cavity; i.e. where it does not vanish it must lie tangentially against the surface: the electric field on the other hand cannot lie along the boundary but where it does not vanish must stand at right angles to the surface.

Evidently, as appears from Fig. 222, the standing-wave pattern of Fig. 221 can be fitted into a rectangular box, formed by placing conducting partitions across the waveguide a distance apart equal to $p\lambda_g/2$ where p is an integer.

It is then possible to fit p cells of the pattern into the box provided the end walls touch but do not cut a magnetic loop. A possible disposition is shown in Fig. 222 for the case $p = 2$. Figs. 222 (a) and (b) illustrate respectively the fields at times $t = 0$, when the field is entirely magnetic, and $t = T/4$, when the field is entirely electric.

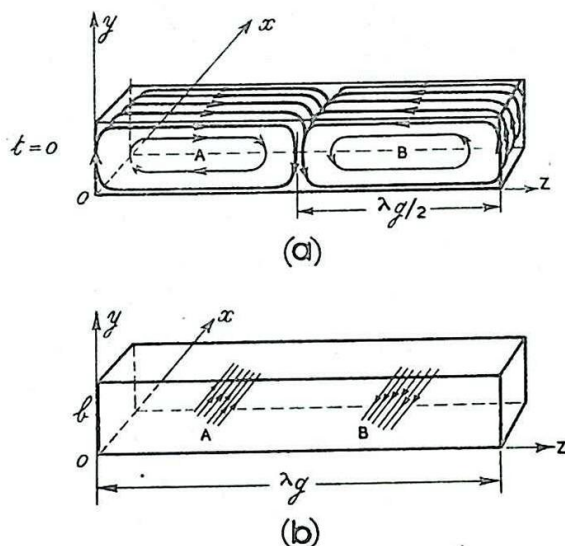


Fig. 222 - Field patterns of the H_{012} mode in a rectangular cavity.

If we choose a set of cartesian axes of reference with the origin 0 at the near left-hand corner, as shown, then with respect to these axes the constituent travelling waves of Fig. 221 that combine to give the stationary wave are H_{01} -waves. Since $p = 2$, i.e., two cells of the pattern are fitted into the resonator, and since the mode is derived from progressive H_{01} -waves, it is designated H_{012} .

It is evident that the same procedure will lead to the field patterns of more general modes of oscillation both in rectangular and in cylindrical resonators. We first find the standing wave pattern in

the unclosed guide corresponding to the general E_{nm} - or H_{nm} - wave. The guide is then converted to a resonator by the introduction of conducting partitions, at positions such that p cells of the pattern are enclosed within the resonator, without violation of the conditions imposed on the behaviour of the electromagnetic field at the cavity surface. The resulting mode is then designated, according to the type of its constituent progressive waves, an E_{mp} - or an H_{mp} - mode.

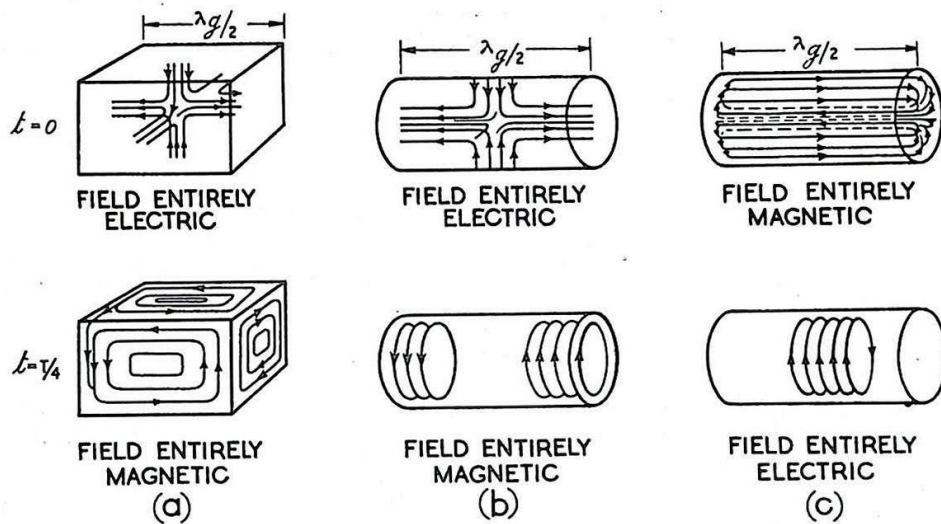


Fig. 223 - Examples of modes of oscillation in cavities.

Examples are shown in Figs. 223 (a), (b) and (c) respectively, of the E_{111} - mode in a rectangular cavity and the E_{011} - and H_{011} - modes in a cylindrical cavity, at times $t = 0$ and $t = T/4$. As shown in the earlier example, the magnetic and electric fields oscillate in quadrature and are displaced relatively by a distance $\lambda_g/4$ in comparison with their positions in the corresponding patterns of the travelling waves.

28. Resonant Wavelength of the Cavity

Since the E_{mp} - and H_{mp} - modes have field patterns the same as those of sets of p elementary cells of the corresponding E_{nm} - and H_{nm} - standing waves in a waveguide, it follows that the length of the resonator must be $p \lambda_g/2$, each cell occupying a length $\lambda_g/2$ of the guide. If, therefore, the length of the resonator is d , resonance occurs for an $n m p$ - mode when

$$d = p \lambda_g/2, \dots\dots\dots (1)$$

where λ_g is the guide - wavelength of the associated E_{nm} - or H_{nm} - wave in a waveguide whose cross section is identical with that of the resonator. Since, the wavelength λ_g is related to the free-space wavelength λ and the cut-off wavelength λ_c by the equation

$$\frac{1}{\lambda^2} = \left(\frac{1}{\lambda_g}\right)^2 + \left(\frac{1}{\lambda_c}\right)^2, \dots\dots\dots (2),$$

it follows from (1) and (2) that the resonant free-space wavelength of the cavity is given by

$$\frac{1}{\lambda^2} = \left(\frac{p}{2d}\right)^2 + \left(\frac{1}{\lambda_0}\right)^2 \dots\dots\dots (3).$$

We consider in turn rectangular and cylindrical resonators.

Rectangular Resonators

The linear dimensions of the resonator shown in Fig. 222 along Ox , Oy and Oz are respectively a , b and d . The cut-off wavelength λ_0 of an E_{rm} - or an H_{rm} - wave in a rectangular guide whose cross section has linear dimensions a and b is given by

$$\left(\frac{1}{\lambda_0}\right)^2 = \left(\frac{n}{2a}\right)^2 + \left(\frac{m}{2b}\right)^2 \dots\dots\dots (4).$$

Finally, the resonant free-space wavelength λ_{rmp} of the E_{rmp} - and H_{rmp} - modes is, from (3) and (4),

$$\left(\frac{1}{\lambda_{rmp}}\right)^2 = \left(\frac{n}{2a}\right)^2 + \left(\frac{m}{2b}\right)^2 + \left(\frac{p}{2d}\right)^2 \dots\dots\dots (5).$$

The resonant frequency is

$$f_{rmp} = \left(\frac{c}{\lambda_{rmp}}\right) = \frac{c}{2} \left\{ \left(\frac{n}{a}\right)^2 + \left(\frac{m}{b}\right)^2 + \left(\frac{p}{d}\right)^2 \right\}^{\frac{1}{2}} \dots\dots\dots (6)$$

where c is the velocity of propagation in free space.

(We have assumed throughout that the cavity is empty
or air-filled)

For example, the resonant frequency of the H_{011} - mode is

$$f_{011} = \frac{c}{2} \left\{ \frac{1}{b^2} + \frac{1}{d^2} \right\}^{\frac{1}{2}},$$

and if $b > a$ this is the lowest frequency at which the cavity can resonate. Since there is no E_{cm} - or E_{no} - wave in a rectangular guide the E - mode of the resonator with lowest frequency is the E_{111} - mode whose frequency f_{111} is also that of the H_{111} -mode,

$$f_{111} = \frac{c}{2} \left\{ \frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{d^2} \right\}^{\frac{1}{2}}.$$

The resonant wavelength of the H_{011} - mode when the resonator is a cube ($a = b = d$) is given by

$$\frac{1}{\lambda_{011}} = \frac{1}{2} \cdot \frac{\sqrt{2}}{a}$$

or $\lambda_{011} = \sqrt{2} a$, the diagonal of a face. This shows, as mentioned in Sec. 25, that the fundamental wavelength is of the order of magnitude of the linear dimensions of the cavity.

The smallest cube that will resonate at a wavelength of 10 cms ($f_{011} = 3,000$ Mc/s) has an edge length of 7.07 cm.

Cylindrical Resonators

The cut-off wavelengths λ_c of waves in circular guides are not given by a simple formula such as (5) but depend on the roots of Bessel functions. The cut-off wavelengths in a waveguide of radius r_g , for some of the lower order modes are :-

Mode	H_{11}	E_{01}	$H_{01} \text{ \& } E_{11}$
λ_c	$3.61 r_g$	$2.61 r_g$	$1.64 r_g$

The resonant wave-lengths of these modes in a cylindrical cavity of radius r_g and length d , are therefore given by

$$\frac{1}{\lambda_{\text{imp}}^2} = \left\{ \begin{array}{l} \left(\frac{1}{3.61 r_g} \right)^2 + \left(\frac{1}{2d} \right)^2 \dots\dots\dots H_{11} \cdot \\ \left(\frac{1}{2.61 r_g} \right)^2 + \left(\frac{1}{2d} \right)^2 \dots\dots\dots E_{01} \dots (7) \cdot \\ \left(\frac{1}{1.64 r_g} \right)^2 + \left(\frac{1}{2d} \right)^2 \dots\dots\dots H_{01} \text{ and } E_{11} \cdot \end{array} \right.$$

29. Charges and Currents on Internal Surface of Resonator

When electric lines of force begin on one portion of the boundary and end on another, they terminate on electric surface charges of opposite sign. For instance, in the resonator shown in Fig. 222 (b) positive charge resides on the region around A and negative charge on that around B, with corresponding compensating charges on the opposite wall. When the magnetic field is present skin currents flow on the interior surface everywhere at right angles to the contiguous tangential magnetic field. Where there is no surface field there is no current. We consider the simple case of the H_{011} -mode in a rectangular resonator at the moment when the field is entirely magnetic (Fig. 224 (a)) and later when it is entirely electric (Fig. 224 (b)). The lines of current flow are shown running perpendicular to the magnetic field at the surface. The current is shown converging towards the central region of face ABCD (Fig. 224 (a)) and away from face A'B'C'D'. These faces become fully charged one quarter of a cycle later, as shown in Fig. 224 (b), and electric lines of forces run from the positive to the negative charges.

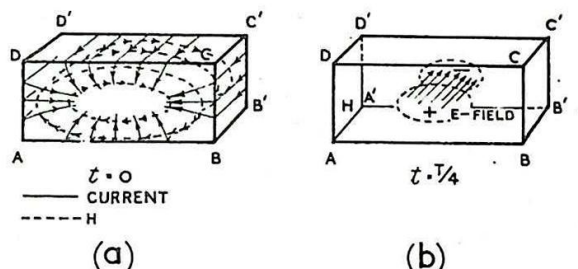


Fig. 224 - Wall currents in a cavity resonator.

30. Methods of Excitation of a Cavity Resonator

A cavity resonator may be excited either by a loop (Fig. 225 (a)) or a probe (Fig. 225 (b)). In the former case the loop must be so introduced that lines of magnetic force can thread through it. Fig. 225 (a) shows two of several possible positions for the loop. As shown, one loop could be an input loop and the other an output loop. The degree of coupling can be controlled by rotation of the plane of the loop.

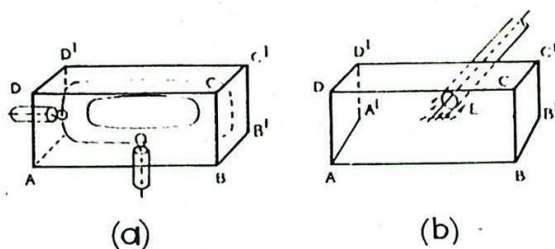


Fig. 225 - Excitation of the H_{011} -mode in a rectangular resonator.

When a probe is used it is introduced into a place of maximum electric field strength and is set parallel to the electric field. Thus in Fig. 199 (b) the probe is shown projecting into the cavity from the face $A'B'C'D'$.

Excitation to resonance can also be made by means of a slot cut in a face such as to interrupt the flow of current. This face could be made common with the wall of a waveguide from which current could be fed into the cavity.

31. The Q-factor of a Cavity

Because of the finite conductivity of the walls, power is dissipated as heat in the walls, and free oscillations decay exponentially. The Q-factor of the cavity is defined by the expression

$$Q = 2\pi \left(\frac{\text{Energy stored}}{\text{Energy dissipated per cycle}} \right) \dots\dots\dots (6).$$

In practice the walls are made of copper or silver and the energy dissipated per cycle is only a small fraction of the stored energy; consequently Q is very large.

Let W be the energy stored; then the energy dissipated per cycle is $\left(\frac{dW}{dt}\right)T$ and expression (6) may be written:

$$Q = \frac{2\pi}{T} \cdot \frac{W}{\left(\frac{dW}{dt}\right)} = \omega \cdot \frac{W}{\left(\frac{dW}{dt}\right)} \dots\dots\dots (7).$$

Whence,

$$\frac{dW}{dt} = \frac{\omega}{Q} W \dots\dots\dots (8).$$

Thus,

$$W = W_0 e^{-\frac{\omega}{Q} t} \dots\dots\dots (9),$$

where W_0 is the stored energy at $t = 0$.

If, therefore, the resonator is shock-excited and left to oscillate freely, the stored energy is reduced to $1/\epsilon$ of its initial value in a time $t' = \frac{Q}{\omega} = \frac{QT}{2\pi}$,

so that $Q = 2\pi \frac{t'}{T}$ (10).

According to (10) an alternative interpretation of Q is that it is 2π times the number of cycles required for the stored energy to decay to $1/\epsilon$ (approximately one-third) of its initial value.

Since Q values of 10^4 and greater are easily achieved, it is evident that a cavity will ring for a great many cycles before the stored energy is reduced to a small fraction of its initial value. A useful approximate rule which gives the order of magnitude of Q in terms of the skin depth δ of the wall currents, and the dimensions of the cavity is given by :-

$$Q \doteq \frac{\text{Volume of Cavity}}{\delta \times \text{surface of Cavity}} \dots\dots\dots (11).$$

A formula for δ in terms of the wave length and wall conductivity is

$$\delta = 2.82 \cdot 10^{-2} \cdot \sqrt{\frac{\lambda}{\sigma}} \text{ metres.}$$

where δ and λ are in metres and σ is in mhos per metre cube.

Suppose the resonator to be made of copper for which $\sigma = 5.8 \times 10^7$ mhos per metre cube. Then $\delta = 3.7 \times 10^{-6} \sqrt{\lambda}$ metres.

According to equation (6) the Q -factor for a cubical resonator of edge a metres, is

$$Q \doteq \frac{a}{6\delta} = \frac{a}{6 \cdot 3.7 \sqrt{\lambda}} \frac{10^6}{\sqrt{\lambda}} \dots\dots\dots (12).$$

For the H_{011} -mode, $\lambda = a\sqrt{2}$, then,

$$Q = \frac{10^6 \sqrt{\lambda/2}}{22 \cdot 2}.$$

For $\lambda = 10$ cm ($= \frac{1}{10}$ metre), this becomes

$$Q = 1.1 \cdot 10^5$$

Such a cavity, if shock-excited, would ring for $\frac{Q}{2\pi} \doteq \frac{10^5}{6} = 1.7 \cdot 10^4$ periods before the stored energy was reduced to $\frac{1}{\epsilon}$ of its initial value.

Each period ($\lambda = 10$ cm) is $\frac{1}{3000}$ microsecond so that the time of ring is about 5 microseconds (or about $\frac{1}{2}$ mile of radar range).

Because they are highly selective, cavity resonators are used as wave metres at centimetre wavelengths. This, and other applications of importance in radar are described below.

32. Applications of Cavity Resonators

(i) Wave Meters

Coaxial Line Wave Meter A convenient wave meter for use at wavelengths of 9-11 centimetres is the coaxial line wavemeter shown in Fig. 226. It comprises a cylindrical cavity into which a rod can be intruded axially to any desired extent by means of a rack and pinion. The metal block serves as a guide for the rod and as a short-circuit to the coaxial transmission line system formed by the cavity and the rod.

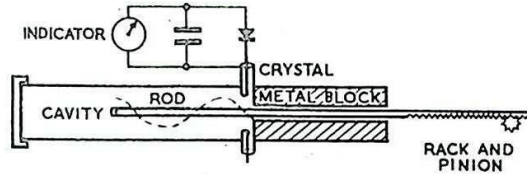


Fig. 226 - Coaxial line wave-meter.

The cavity is excited by injecting an EMF into the input loop from the source whose wavelength is required, and the rod is moved by means of the pinion until

a position of resonance is indicated by the detecting crystal and microammeter fed from the output loop. The shortest resonant length l of the rod within the cavity is slightly less than $\lambda/4$ since the transmission line system is open-circuited but with some small end-capacitance. Other resonant positions correspond to $(l + n\lambda/2)$ where n is an integer, since the end capacitance remains the same independent of the position of the end of the rod. This is because the end of the rod never closely approaches the closed end of the cavity.

The rack and pinion carry a scale and vernier from which the displacement of the rod can be measured directly in centimetres. The displacement of the rod between successive resonances is equal to $\lambda/2$ and gives directly the wavelength on the coaxial line, which is the same as the free-space wavelength of the source.

The diameter of the cavity is small enough to ensure that hollow cavity modes of oscillation cannot occur.

Resonant Cavity Wave Meter A wavemeter suitable for measuring changes in wavelength with great accuracy is the cavity wavemeter shown in Fig. 227 (a). It is a metal cylinder whose length can be adjusted by rotation of the screwhead to which the upper end of the cylinder is attached. The resonant mode employed is the H_{011} -mode which is excited by an input loop near the middle of the cavity, as shown, and resonance is indicated by means of an output loop, at the same height but shifted through a quadrant, and a crystal detector and microammeter combination. The magnetic lines of force at resonance are indicated in Fig. 227 (a). To prevent the excitation of the E_{11} -mode, which has the same resonant frequency as the H_{011} -mode, the movable end clears the cylinder wall with a small gap. The currents in the H_{011} -mode flow on the cavity surface in circles about the axis (Fig. 227) and no flow occurs from the flat ends to the curved surface. The current distribution is therefore indifferent to the presence of the gap between the movable flat end and the curved wall. In all other modes, except H_{0nm} -modes, current is required to flow from the flat ends to the curved wall; consequently the introduction of the gap effectively suppresses these unwanted modes. Wire filters of suitable form can also be used as suppressors but they are less convenient. In principle the resonant wavelength could be obtained from the dimensions of the cavity at resonance, but in practice it is more

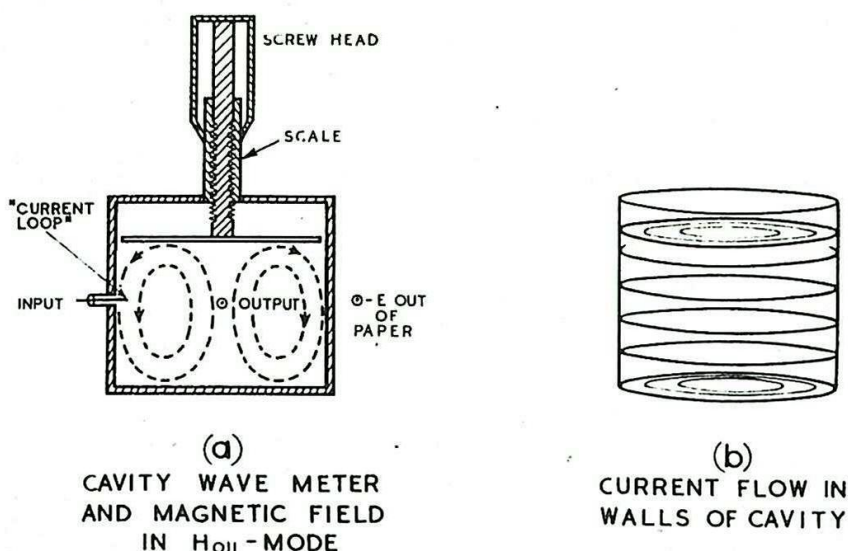


Fig. 227 - Cavity wave-meter and magnetic field in H_{011} -mode (a). Current flow in walls of cavity (b).

accurate to measure displacements of the lid by means of a micrometer screw and scale, as indicated. This scale is then calibrated against the harmonics of a crystal oscillator, or, more crudely, against the coaxial line wavemeter described above, using a tunable source of EMF. The wavelength scale is very open and a small change in wavelength corresponds to a relatively large displacement of the movable end. The interior of the cavity is usually silvered so that a large Q and sharp resonance results. The wavemeter is then suitable for examining the RF spectrum of a magnetron pulse (Fig. 228).

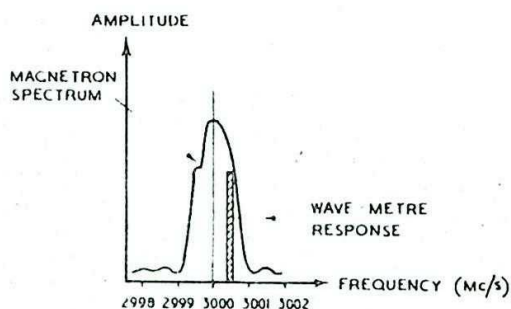


Fig. 228 - Use of wave-meter for examining magnetron frequency spectrum.

(ii) Echo Boxes

It is often necessary to check the overall performance of a centimetre wave radar equipment in situations where it is difficult to obtain echoes from objects at suitable ranges. For instance a radar equipment in an aircraft may have its scanner directed downwards so that it is impossible to obtain echoes when the aircraft is on the ground. The echo box is a simple device for checking roughly the overall performance of a set. It is merely a resonant cavity designed to possess a high Q . The cavity is shock-excited by the transmitter pulse and continues to ring and emit a signal which is spread along the time base for an appreciable distance after the cessation of the transmitter pulse. A possible arrangement is shown in Fig. 229. The echo box is fed via a screened low-loss cable from a pick-up probe

fixed near the edge of the mirror. The energy abstracted from the transmitted pulse is stored as a resonant mode in the box and is re-radiated to the receiver as an exponentially decaying signal. For the greater part of its duration the signal saturates the receiver but finally decays to a level at which it no longer does so. The appearance on a Type A display is illustrated in the figure. The range at which the echo box response disappears into the noise gives an indication of the overall performance of the set.

The energy abstracted from the

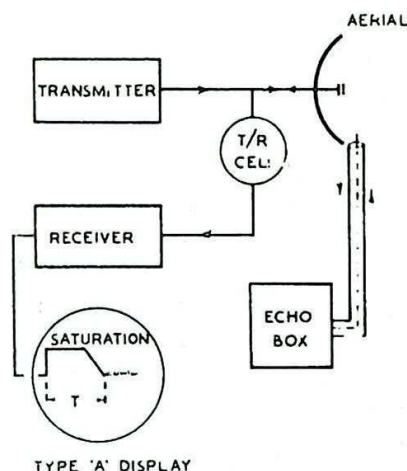


Fig. 229 - Arrangement of radar system with echo box.

Echo boxes are of two types, tuned and untuned. Tuned echo boxes are the same as the wave meter already described and illustrated in Fig. 227. They require tuning to the frequency of the transmitter which then shock-excites the H_{011} -mode. The output loop is not used.

Untuned echo boxes are very large cavity resonators whose lowest modes, H_{111} , E_{111} etc., correspond to wavelengths much greater than the wavelength of the radar set. They are usually hollow cubes with copper walls.

The modes that are excited by the radar transmitter are therefore higher order E_{mnp} - and H_{mnp} -modes where m , n and p are relatively large integers.

Consider a cubical resonator whose edge a comprises several wavelengths.

It can be shown that the spacing of the higher modes is such that the number of modes comprised within the wavelength range λ to $(\lambda + \Delta\lambda)$ or frequency range f to $(f + \Delta f)$ in a rectangular resonator, is

$$N \doteq 8\pi \frac{V}{\lambda^3} \cdot \left(\frac{\Delta\lambda}{\lambda}\right) = 8\pi \frac{V}{\lambda^3} \cdot \left(\frac{\Delta f}{f}\right)$$

where V is the volume of the resonator. Thus, for a cubical resonator of edge a

$$N \doteq 8\pi \left(\frac{a}{\lambda}\right)^3 \left(\frac{\Delta\lambda}{\lambda}\right) = 8\pi \left(\frac{a}{\lambda}\right)^3 \left(\frac{\Delta f}{f}\right).$$

Consider the case of a resonator of edge $a = 1$ metre excited by a magnetron pulse for which $f = 3000$ Mc/s, ($\lambda = 10$ centimetres) and the bandwidth is 1 Mc/s. The formula gives, for the number of modes covered by the bandwidth,

$$N \doteq \frac{8\pi \cdot 10^3}{3 \cdot 10^5} = \frac{8\pi}{3}$$

i.e. there are 8 modes. Thus, the transmitter pulse is able to

excite a number of resonant modes whatever its frequency f and it is not necessary to tune the resonator. In the case discussed, the mean frequency separation of the modes is $\frac{1}{8}$ Mc/s = 125 kc/s.

Since the ratio of volume to surface of a cubical resonator is proportional to the edge a , these large untuned resonators possess very high Q values and will ring for many microseconds.

(iii) Cavity Resonators in Centimetre Wave Oscillators

The usual resonant circuits, comprising lumped inductances, capacitances and resistances, cannot be constructed in a useful form at wavelengths of 10 centimetres and less, and it is necessary to replace them by other resonant systems such as resonant lengths of coaxial transmission line or resonant cavities. In the klystron oscillator, whose main application is as low-power local oscillator in centimetre wave radar receivers, a resonant cavity is used in which a mode of oscillation is maintained by a bunched electron stream. In order to bring about bunching and to abstract power from the bunched stream, the electrons must pass through the oscillating fields within the cavity against or parallel to the electric field where it is most intense. Further, the time of transit must be small compared with the period of oscillation. It is not possible to accomplish this in any of the resonators described so far because their dimensions are comparable with the wavelength λ and it would be necessary for the electrons to move at speeds comparable with the velocity of light in order that the transit time should be less than the period of oscillation. What is required therefore is a resonant cavity (rhumbatron) in which an intense oscillating electric field is concentrated across a short path so that electrons can travel the whole extent of a line of force in a time short compared with the period.

In one type of reflector klystron, the Sutton Tube, an example of which is the CV67, the resonator (rhumbatron) assumes the form indicated in Fig. 230 (b). This rhumbatron may itself be regarded as a distorted form of a the prototype shown in Fig. 230 (a). In Fig. 230 (a), A and B represent a pair of coaxial conducting cones with their tips removed so as to form a small gap between them. It is known that when an alternating EMF is applied across the gap the pair of cones forms a transmission line system and that a principal (TEM) wave is guided along them. The lines of electric and magnetic force run as shown in the figure. If a conducting spherical surface C of radius $\lambda/4$, concentric with the middle of the gap, is used to close the cones then the TEM wave is reflected without distortion and a complete standing wave is produced on the transmission line which then forms a resonant system. The equivalent twin transmission line system is shown in Fig. 230 (c). We may, however, regard the system of Fig. 230 (a) as a hollow spherical cavity resonator with a pair of conical projections. It follows from what has been said that the fundamental resonant wavelength of this resonator is four times the radius of the sphere. A voltage antinode is located at the gap AB where the electric field attains its greatest intensity. It is possible therefore to maintain such a rhumbatron in resonance by passing a bunched electron stream across the gap AB. In practice, in the reflector klystron used as a local oscillator, it is necessary to control the resonant frequency by means of external tuning screws; consequently, the rhumbatron is divided in two by a glass tubular envelope which is evacuated and contains the electron gun assembly and the reflecting electrode. The portion of the resonator external to the glass envelope carries the tuning screws. To introduce the glass envelope it is necessary to distort the shape of the rhumbatron to that

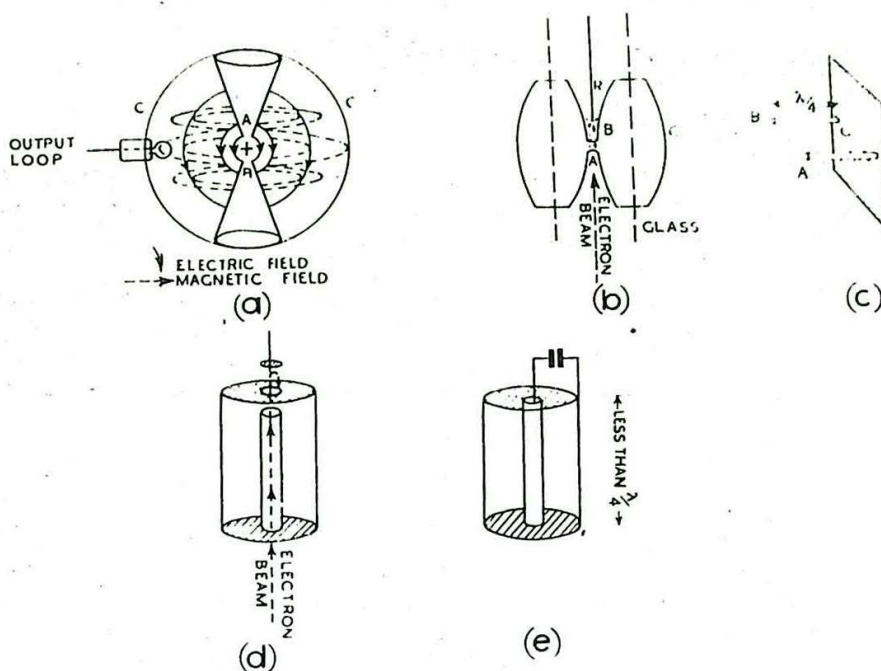


Fig. 230 - Common types of resonators.

shown in Fig. 230 (b). One of the cones is also distorted to bring the reflector close to the gap. Power is abstracted through a loop placed, as shown in Fig. 230 (a) with its plane parallel to the axis of the cones.

The input impedance to the transmission line system shown in Fig. 230 (a) is large at resonance, and it is necessary to drive the rhumbatron from a high impedance source. Thus the power supplied by the electron stream must be in the form of relatively high voltage electrons and relatively small current. Thus, in the OV67, the accelerating voltage is 1200 volts and the electron current is 6 milliamps. The maximum power output at $\lambda = 9$ centimetres is 200 milliwatts which is ample for a local oscillator.

The high operating voltage is an inconvenience and a more convenient form of klystron operates on a voltage of 300 and is therefore able to use the same power pack as the radar receiver. The resonator here comprises a short-circuited coaxial transmission line with a gap between the inner and the end plate of the outer at the top, as shown in Fig. 230 (d). This end plate has a hole in its centre, and both this hole and the end of the inner are covered by a wire gauze through which the electron stream passes into and out of the gap. The length of the transmission line is less than $\lambda/4$ and it is brought to resonance by the capacitance between the free end of the inner conductor and the end plate.

The equivalent resonant system is shown in Fig. 230 (e). The shunting impedance of this system is much lower than that of the double cone system and the power carried by the electron stream can be supplied at lower voltage and larger current.

Various types of cavity resonators which may be employed in the Resonant Cavity Magnetron are illustrated in Chap. 8 Secs. 27 and 31, where the many possible modes of oscillation are discussed.

CHOICE OF WAVEGUIDE SHAPE AND DIMENSIONS

33. General Considerations

In foregoing sections we have reviewed the principal features of electromagnetic disturbances in metal tubes. It remains to describe how this basic knowledge is applied to useful ends.

The primary function of a waveguide is to convey electromagnetic power from a generator to an aerial or from an aerial to a receiver, at wavelengths of ten centimetres or less. In fact, waveguides are required to perform the same functions at these wavelengths as are performed by transmission lines at longer wavelengths.

The principle problems of transmission line practice are, to produce a reflectionless termination (matching), to achieve aerial switching and a satisfactory system of common aerial working (common T/R) and to measure standing wave ratios (impedance measurements). To achieve these objectives certain ancillary devices such as reactive stubs, quarter-wave insulators, quarter- and half-wave transformers and the like, are employed. In what follows we shall discuss the parallel problems of waveguide practice and investigate the character of the analogues of the ancillary devices.

When waveguides were first recognised as providing the most suitable method of conveying large powers from a source to an aerial at wavelengths of 10 cms and less, it became a matter of practical importance to answer correctly the following questions :-

- (i) what is the most suitable size for a waveguide operating on a specified wavelength ?
- (ii) what is the best geometrical form to give it ?

We shall consider these queries in the following sections.

34. Choice of Waveguide Dimensions

Transmission line practice is based on the fact that at the usual frequencies progressive waves exist only in the principal or TEM mode, all other modes being evanescent. In order to adapt transmission line practice to waveguides it is therefore necessary to ensure that progressive waves are present in one mode alone, all other modes in the waveguide being evanescent. This result is achieved by so choosing the cross-sectional dimensions of the guide that the free-space wavelength of the wave to be propagated is less than the greatest of the cut-off wavelengths of the various modes in the waveguide, but greater than those of all other modes. In this way the mode with the greatest cut-off wavelength is present as a progressive wave, but all other modes are evanescent.

For instance, in the example given in Sec. 13, the longer cross-sectional dimension of a rectangular waveguide designed to carry waves whose free-space wavelength λ is 9 cms, is $b = 2\frac{1}{2}$ inches. The cut-off wavelength of this guide for the H_{01} -mode is $\lambda_c = 2b = 5$ inches = 12.7 cm. The H_{01} -wave is therefore propagated in this waveguide, but as the cut-off wavelengths of all other E_{TM} - or H_{TM} -modes, given by :-

$$\frac{1}{\lambda_c^2} = \left(\frac{n}{2a}\right)^2 + \left(\frac{m}{2b}\right)^2$$

when $a < b$, are less than $\lambda = 9$ cms, these modes are evanescent.

35. Geometrical Form of Waveguide Section

The choice, in practice, lies between circular and rectangular sections and it is found that, except in specialised devices such as rotary joints, circular sections are unsatisfactory.

For instance, the wave in the circular waveguide would normally be an H_{11} -wave since this mode possesses the greatest cut-off wavelength. Should, the tube however, be slightly elliptical over an appreciable length the H_{11} -wave would then resolve into two waves with different values of u_p , so that on recombination where the tube again assumes a circular section, the plane of polarisation in the resulting H_{11} -wave is rotated with respect to the original plane. This difficulty of preserving the polarisation of the field pattern over a long run of waveguide is a disadvantage of circular guides. Consequently, when circular waveguides are employed they appear only in short and straight lengths, and in practice long runs of waveguide are almost invariably rectangular in cross section.

The dimensions a and b of the rectangular cross section are always made unequal in such guides. The smaller dimension a is such that $2a$ is less than the wavelength of the H_{01} -mode, but the dimension b accepts the H_{01} -mode as a progressive wave but no other modes. Thus the field pattern preserves a unique sense and can be propagated only with the transverse electric field parallel to the shorter edge a . The following are dimensions of typical waveguides:-

<u>Free-space wavelength</u>	<u>Internal Guide Dimensions</u>
10 cm.	3" x 1"
9 cm.	2½" x 1"
3.3 cm.	1" x ½" (British)
	0.9" x 0.4" (American)

It is undesirable to operate the waveguide near the cut-off frequency of the H_{01} -wave since dispersion (that is, the dependence of u_p on frequency) is very marked near the cut-off frequency (Sec. 6). Also there is a marked increase in attenuation as the cut-off frequency is approached (Sec. 23).

In a long run of waveguide dispersion can affect the shape of a pulse. The dimensions given in the table above remove the cut-off wavelength well away from the operating wavelength; for instance, in the 9 cm. guide whose dimensions are 2½" x 1", the angle of incidence ($90^\circ - \alpha^\circ$) (Fig. 204) of the waves successively reflected between the walls is 45° , whereas at cut-off it is zero.

STANDING WAVES IN WAVEGUIDES

36. Waveguide Impedances

It is shown in Chap. 17 Sec. 7 that when a plane-polarised electromagnetic plane wave is propagated in free space, the wave front

is presented with the equivalent of a resistance whose magnitude is given by

$$R_0 = 120 \pi \text{ ohms.}$$

This is called the Wave- or Field-impedance. If E is measured in volts per metre and H in amps. per metre then this may be written

$$\frac{E}{H} = R_0 = 120 \pi \text{ ohms.}$$

A similar result holds for waves propagated in waveguides. Considering the H_{on} wave of Sec. 7, depicted in Figs. 199 and 200, there are seen to be two components of the H -field, one along the guide and the other transverse (vertical) whilst there is a single transverse component of the E -field perpendicular to the plane of the paper. If axes x , y , and z , are taken as indicated in Fig. 195 these components are given by:-

$$H_y = 2H \cos \alpha \sin\left(\frac{2\pi y}{\lambda} \cdot \sin \alpha\right) \cos\left(\omega t - \frac{2\pi z}{\lambda} \cos \alpha\right)$$

$$H_z = 2H \sin \alpha \cos\left(\frac{2\pi y}{\lambda} \cdot \sin \alpha\right) \sin\left(\omega t - \frac{2\pi z}{\lambda} \cos \alpha\right)$$

$$E_x = 2E \cdot \sin\left(\frac{2\pi y}{\lambda} \cdot \sin \alpha\right) \cos\left(\omega t - \frac{2\pi z}{\lambda} \cos \alpha\right)$$

H , E , are the amplitudes of the H - and E -fields of the component waves; α is the angle between the direction of propagation of the component waves and the reflecting plates (Fig. 200).

Whereas H_y and E_x oscillate in phase, H_z is in quadrature with the other two, and may be considered as supplying the reactive or storage component of the resultant fields in the guide. The other two components convey the energy along the guide. By division we obtain

$$\frac{E_x}{H_y} = \frac{E}{H} \sec \alpha.$$

Since $\frac{E}{H}$ is the ratio of E -field to H -field in a plane polarised electromagnetic plane wave in free space, of the type considered in Chap. 17 Sec. 7, its value is $120\sqrt{\pi} \doteq 377$ ohms.

Hence

$$\frac{E_x}{H_y} = 120\sqrt{\pi} \sec \alpha \doteq 377 \sec \alpha \text{ ohms.}$$

Since, for an H_{on} -mode, $\sec \alpha = \frac{\lambda_g}{\lambda}$, this equation may be written

$$\frac{E_x}{H_y} = 120\sqrt{\pi} \frac{\lambda_g}{\lambda}.$$

In general the equation $\frac{E \text{ transverse}}{H \text{ transverse}} = 120\sqrt{\pi} \frac{\lambda_g}{\lambda}$ holds for all

H_{nm} -waves propagated in rectangular guides. The corresponding result for E_{nm} -waves is

$$\frac{E \text{ transverse}}{H \text{ transverse}} = 120\sqrt{\pi} \frac{\lambda}{\lambda_g}$$

(In this case it is the E -field which is resolved into two components, whilst the H -field is entirely transverse, so that the effective ratio $\frac{E}{H}$ is less than that for free space).

This ratio $\frac{E \text{ transverse}}{H \text{ transverse}}$ is called the Wave Impedance

for the mode considered, and has a uniform value at all points of the wave front of a progressive wave.

If a waveguide is terminated in a resistive film whose resistance per unit area is equal to the wave impedance the guide is properly matched for the mode considered (see Sec. 37 and 38). For any other termination some reflection occurs and standing waves are developed.

Whereas it is not usually practicable to measure the effective impedance of the termination, the standing wave ratio may be determined by means of a standing wave indicator, and various means may be adopted of eliminating the standing waves on the main run of the guide. Since for a single mode in a given guide the SWR and the position of nodes and antinodes gives a sufficiently complete description of the standing wave pattern for practical purposes, it is possible to apply the same techniques for dealing with standing wave problems in waveguides as are used for transmission lines. In particular, circle diagrams may be used, the value of the normalised impedance indicating the position in the guide relative to the nearest E-node or antinode for a given SWR. Since we are not interested in the actual impedances in the guide it is not necessary to use any but normalised impedances and admittances, for which the same symbols will be used as for transmission lines.

Where circle diagrams are used to illustrate waveguide properties it is assumed that the reader is familiar with the treatment of circle diagrams given in Chap. 4.

37. Reflection from the Waveguide Termination

As in the case of transmission lines, unless a waveguide is properly terminated, energy will be reflected and standing waves will arise. The ideal termination must have an impedance equal to the characteristic impedance of the guide for the mode considered. If the normalised impedance presented by the termination is $[z_p]$ the fraction of the wave reflected is given by :-

$$\rho = \frac{[z_p] - 1}{[z_p] + 1}.$$

ρ is called the reflection coefficient (compare Transmission Lines, Chap. 4, Sec. 9). If $[z_p] = 1$ the waveguide is properly matched and no reflection occurs. If the guide is terminated in a perfectly conducting metal sheet, $[z_p] = 0$ and $\rho = -1$, so that the termination acts as a short-circuit and at the termination the E-field of the reflected wave is in opposition to the E-field of the direct wave. If $[z_p] = \infty$, the termination acts as an open-circuit, and at the termination the H-field of the reflected wave is in opposition to the H-field of the direct wave. In waveguides, however, if the end of the guide is left open, the terminating impedance is not infinite but possesses a resistive component, due to the radiation from the end of the guide, and a reactive component due to evanescent modes excited at the discontinuity; (see Sec. 42).

The equivalent of an open-circuit termination is provided by a short-circuited $\frac{\lambda_g}{4}$ section of the guide (compare Chap. 4, Sec. 26). The input impedance of such a section, if losses are neglected, is infinite.

A theoretically perfect match is obtained for any mode if the guide is terminated in a resistive film whose resistance per unit area is equal to the wave-impedance for that particular mode in the guide. The film must be backed by a short-circuited $\frac{\lambda_g}{4}$ -section of the guide so that the finite impedance at the open end does not appear in parallel with the terminating impedance. Other methods of correctly terminating a waveguide are discussed in the next section.

Where the guide is improperly terminated, so that a standing wave occurs, the standing wave ratio is given by :-

$$S = \frac{\hat{E}}{\check{E}} = \frac{\hat{H}}{\check{H}} = \frac{1 + |\rho|}{1 - |\rho|}$$

where \hat{E} , \check{E} and \hat{H} , \check{H} are maximum and minimum values of the transverse components of electric and magnetic fields respectively, measured at corresponding positions of the guide cross-section.

38. Practical Methods of Achieving a Reflectionless Termination

In order to test a waveguide run for reflections at joints and bends and to measure the impedance of obstacles, stubs and slots, it is necessary to be able to terminate the waveguide in a reflectionless termination. Any reflected wave then appearing in the guide must be due to some cause other than the termination.

A practical form of termination that gives negligible reflection and is suitable for work at low powers is the wooden load. Wood, as a material, heavily absorbs electromagnetic radiations at centimetre wavelengths; consequently if a wooden wedge of either of the forms of Fig. 231 is inserted into the end of a rectangular waveguide which it fits closely, the two component plane waves of the H_{01} -wave in the guide enter the wood at the sloping face and penetrate into its interior where they are absorbed so that there is little reflection at the surface.

The sloping faces of the loads are parallel to the electric vector of the incident H_{01} -wave, consequently Fig. 231 represents a view of the broad face of the waveguide. That portion of the wave which is reflected from the wood surface continues to travel along the guide in the same direction, obeying the optical law

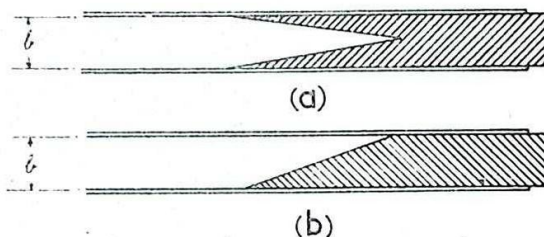


Fig. 231 - Reflectionless terminations.

Angle of Incidence = Angle of Reflection,

and undergoes successive reflections until the energy is almost completely absorbed. To avoid appreciable reflection the sloping faces should not be less than $2\lambda_g$ in length; the SWR in the guide is then of the order 1.05:1.

In a circular guide the reflectionless load is a long cone with a cylindrical butt-end. As an alternative to wood, a composition of bakelite and graphite is sometimes used, and for work at high powers,

where the wood could char, a mixture of graphite and sand is employed.

In both rectangular and circular guides there is an optimum value for the angle between the waveguide walls and the sloping surface of the load.

39. Attenuating Section

In measurements of standing wave ratios, in order to determine the reflection coefficients and impedances of waveguide elements and loads it is essential to isolate the generator from the standing wave indicator and the termination in order to avoid variations in the load on the generator which could affect both its frequency and power output. This is achieved by the inclusion of an attenuating length of waveguide which is merely a length of normal rectangular guide with a thin strip of absorbing material, wood or bakelite impregnated with graphite, fixed internally to one of the narrow faces. The attenuation of some 10 db. in the signal which is thus achieved is sufficient to isolate the generator. No reflected wave of any appreciable amplitude then reaches it.

40. Reflection from an Irregularity in a Waveguide

If an obstacle or irregularity is introduced into a waveguide, electromagnetic energy is reradiated or "scattered" from it so that a wave is reflected back along the guide (Fig. 232). Another wave is generally reradiated from the obstacle away from the source, but this does not interfere with the input admittance of the guide and is, for the moment, ignored. The obstacle may thus be regarded as introducing an additional impedance or admittance either in series or in parallel with the impedance already present. Whether this impedance must be considered as a series or as a shunt impedance depends on the shape and position of the obstacle in relation to the fields and walls of the guide.

A theoretical investigation, into which we shall not enter here, leads to the conclusion that when the discontinuity scatters symmetrically (that is, the transverse electric vectors in the wave fronts of the two waves scattered in opposite directions are equal and similarly directed at equal distances from the scattering section) then the discontinuity behaves like an impedance (or admittance) placed in shunt across a transmission line. If, however, the scattering is anti-symmetrical (amplitudes at equal distances are equal but the electric vectors are oppositely directed) then a series representation is required.

Obstacles possessing geometrical symmetry when placed with an axis of symmetry parallel to the electric vector, scatter symmetrically and therefore behave as shunt elements.

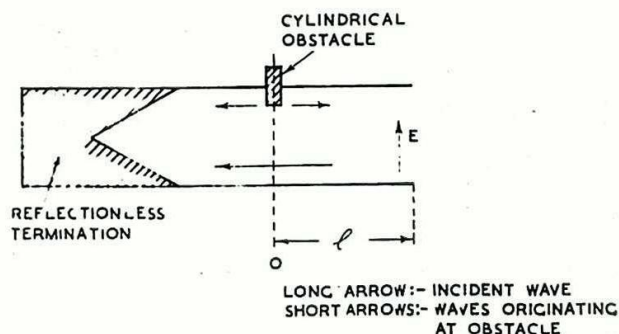


Fig. 232 - Scattering of waves by an obstacle in a rectangular guide.

Determination of the Admittance of an Obstacle

Consider a rectangular waveguide (Fig. 232) terminated in a reflectionless load, within which a metallic obstacle has been placed. The obstacle may be, for instance, a thin metal cylinder that projects a short distance into the waveguide, with its axis parallel to the short edge of the cross-section. When the cylinder is withdrawn the electromagnetic field in the guide is merely that of the H_{01} -wave travelling from the generator to the load.

When the obstacle is introduced at the dotted section it distorts the field of the H_{01} -wave in its vicinity because the component of the electric field tangential to the surface of the obstacle must vanish. In the particular case of the cylinder the electric lines of the H_{01} -wave which themselves run parallel to the short edge of the cross-section are also parallel to the axis of the cylinder. The distorted field near the dotted section can be resolved into the following constituents :-

- (i) The original H_{01} -incident wave (indicated by the long arrow in Fig. 232).
- (ii) A pair of H_{01} -waves originating at the obstacle, the one propagated towards the generator and the other towards the load (these waves are indicated by short arrows in Fig. 232).
- (iii) A series of evanescent modes which are prominent near the obstacle but disappear at a sufficient distance from it. These form the storage field.

The relative amplitudes and phases of the components (i), (ii) and (iii), of the field are such that the tangential electric field is zero both over the surface of the obstacle and at the walls of the waveguide. Take the cross-section of the waveguide at which the cylinder is introduced as the section $l = 0$. At another section at a sufficient distance l away towards the generator, the evanescent waves will have become unimportant and the electromagnetic field reduces to that of the incident wave, $A \cos (\omega t + kl)$ with that of the reflected wave $B \cos (\omega t - kl + \phi)$ superimposed. The resultant field E is given by the equation

$$E = A \cos (\omega t + kl) + B \cos (\omega t - kl + \phi),$$

$$= A \left[\cos (\omega t + kl) + |h| \cos (\omega t - kl + \phi) \right],$$

where, $B/A = |h|$ and $k = 2\pi / \lambda_g$. Or in complex notation,

$$E = A e^{j\omega t} (e^{jkl} + h e^{-jkl}),$$

where $h = |h| e^{j\phi}$.

Thus from the standpoint of the section at position l , the section at $l = 0$ possesses a reflection coefficient $h = |h| e^{j\phi}$, when the obstacle is in position, and an associated admittance $y_t = (1 - h)/(1 + h) = G_t + jB_t$.

When the obstacle is removed the reflection coefficient of the section $l = 0$ becomes zero, and its normalised admittance becomes unity.

The coefficient $h = |h| e^{j\phi}$ is the reflection coefficient of the obstacle for a single progressive wave incident on it.

To find the individual admittance y_1 of an obstacle of arbitrary shape it would be necessary to measure it experimentally. The most direct method is to terminate the waveguide in a metal plate and to introduce the obstacle at an E anti-node in the standing wave pattern. At this place the admittance y_t of the section is zero and becomes $y_t = y_1$ when the obstacle is introduced. The standing wave pattern between the obstacle and the generator is now investigated with a standing wave indicator and the admittance y_1 determined from it.

Alternatively, the waveguide can be terminated in a reflectionless termination and the obstacle may be introduced at any convenient position. The impedance of the section then becomes $y_t = 1 + y_1$.

Obstacles that do not themselves absorb power possess admittances y_1 that are pure susceptances, $y_1 = \pm jB_1$. When the susceptance is positive it is termed capacitive and when negative, inductive.

4.1. Standing Wave Indicator

The deleterious effects of standing waves in transmission lines have been discussed in Chap. 4, Sec. 32. Whilst standing waves may be introduced deliberately into short sections of waveguide for matching purposes, their presence in general is equally undesirable, and for similar reasons, as in the case of transmission lines. It is therefore usual to incorporate in the guide some device for detecting the presence of standing waves so that matching systems can be adjusted to minimise them.

As in coaxial lines, the measurement of the standing wave ratio in a waveguide normally necessitates the insertion of a probe or loop into a slot or series of holes in the guide. Alternatively apertures may be made in the guide so that small, evanescent waves are formed outside the guide, indicative of the field strength inside. Care must be taken that the energy radiated through the slot or aperture is negligible, and that the irregularity introduced into the waveguide does not appreciably disturb the mode of propagation (see Sec. 15). The field strength is indicated by a suitable form of galvanometer, such as a neon lamp, valve voltmeter or crystal detector and microammeter. The variation in field strength at different points along the guide indicates the presence of standing waves, and the ratio of maximum to minimum amplitudes gives the standing wave ratio.

If suitably designed, the standing wave indicator may be calibrated so that it fulfils the dual role of measuring both standing wave ratio and also absolute field strength.

A novel form of standing wave indicator, suitable for operation at wavelengths of the order of three centimetres, is illustrated in Fig. 233.

A straight length of rectangular waveguide has two narrow transverse slots cut in the broad face $\lambda_g/4$ apart. A piece of curved waveguide is fixed to the straight waveguide so that the slots are common to both guides where the walls are in contact. Thus power can be radiated from the straight waveguide through the slots into the arms A and B of the curved guide. By this arrangement the radiations from the slots which are excited by the direct wave in the main guide (travelling from left to right in Fig. 233) reinforce to give a wave in the arm B, but cancel to give no radiation in the arm A (compare End-fire Array, Chap. 17, Sec. 39). Conversely, the reflected wave in the main guide causes a wave to travel in A but not in B.

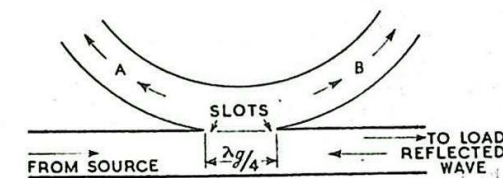


Fig. 233 - Fixed frequency standing wave indicator.

Thus, waves travel in the arms A and B whose amplitudes are proportional to the reflected and incident waves in the main waveguide. In principle, when A and B are terminated in similar field-strength meters the reflection coefficient of the load in the guide may be obtained directly. In practice it is difficult to obtain meters of equal sensitivity and a better arrangement is to terminate B in a reflectionless load (Sec. 38) and to calibrate the output from A against standard mismatches in the main guide. Then, provided the power in the direct wave does not vary, the standing-wave ratio may be read directly from the meter at A. This device can be used at only one wavelength since the slots must be $\lambda_g/4$ apart.

42. Elimination of a Reflected Wave

To eliminate a reflected wave the standing wave pattern is first found by means of a standing wave indicator, and the section at which the admittance assumes the form $[y_t] = 1 + j[B]$ is located from a circle diagram. If now, an obstacle whose normalised susceptance is $[B_1] = -[B]$ is introduced at this section the admittance of the section becomes $[y_t] = 1 + j[B] - j[B] = 1$. (It is assumed that the obstacle may be represented as a shunt impedance (see Sec. 40)). The reflection coefficient $\rho_t = (1 - [y_t]) / (1 + [y_t])$ therefore becomes zero and no reflected wave returns to the generator.

Fig. 234 indicates how the correct position for the obstacle may be found relative to that of an E antinode in the standing wave. An E antinode is a position of minimum admittance \bar{Y} . The point K on the circle diagram corresponds to this position along the guide. The point L corresponds to the position where the admittance is $[y] = 1 + j[B]$. It is a distance $l_1 = n_1 \lambda_g$ away from the E antinode in the direction of the generator where n_1 is the n-value of the arc ML on the circle diagram.

It is required to add the susceptance $-j[B]$ at this point in order to match the waveguide. Alternatively an admittance $+j[B]$ could be added at a distance $l_2 = n_2 \lambda_g$ which corresponds to the point N on the diagram with n_2 as the n -value of the arc MN.

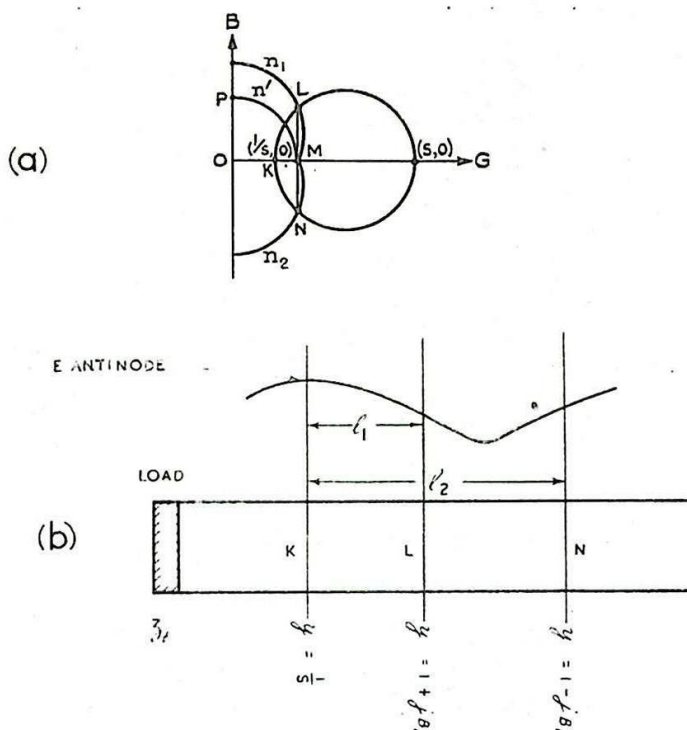


Fig. 234 - Elimination of reflected wave.

It is useful to illustrate what has been said above by actual experimental results, obtained using a standing wave indicator.

A $2\frac{1}{2}$ " x 1" rectangular waveguide with an open end was fed by a 9 cm generator under the following conditions :-

Wavelength λ_g in guide = 13.8 cm (twice distance between adjacent minima).

Standing wave ratio $S = 2.83$.

Distance l_1 of nearest E antinode from the open end

$$= 6.6 \text{ cm} = 0.478 \lambda_g.$$

From the circle diagram we find:-

$$[Z_t] = 2.46 - 0.89j \text{ or } [Y_t] = 0.36 + 0.13j.$$

Thus the impedance $[Z_t]$ is, clearly, not infinite but a resistance and a capacitance in series, or, regarded as an admittance, a conductance and a capacitive susceptance in shunt.

The equivalent circuit representation of the waveguide and its termination $[Z_t]$ is a transmission line of characteristic impedance R_0 terminated in a load comprising a resistance $2.46 R_0$ and a reactance (condenser) $= j0.89 R_0$ in series, or alternatively in a resistance whose conductance is $0.36/R_0$ and a condenser whose susceptance is $j0.13/R_0$ in shunt.

These loads possess the same reflection coefficient and produce the same standing wave pattern on the line as does the open end in the waveguide.

We suppose next that it is required to eliminate the reflected wave in the major portion of the waveguide. We have :-

$$\text{Standing wave ratio } S = 2.83 = \frac{1}{0.354}$$

The admittance at an E antinode is $\frac{1}{S} = 0.354$ consequently, OK in Fig. 234 (a) is equal to 0.354. From a circle diagram it is found that, for the point L, $\boxed{y} = (1 + j\boxed{B})$, $n_1 = 0.163$ and $ML = 1.07$.

We conclude that at a section which lies at a distance $l_1 = n_1 \lambda_g = 0.163 \times 13.8 = 2.25$ cm away on the generator side of any E antinode, the admittance of the guide is there equal to $\boxed{y} = (1 + j.1.07)$. At the point N we find $n_2 = 0.337$ and $MN = -j.1.07$.

The admittance at a distance $l_2 = 0.337 \times 13.8 = 4.65$ cm is therefore $\boxed{y} = (1 - j.1.07)$.

It was found that a complete elimination of the reflected wave could be achieved by protruding a metal screw into the guide through a longitudinal slot in the broad face at the position N (Fig. 234 (b)). The screw was mounted in a holder so that by turning it slowly the extent to which it projected into the guide parallel to the electric field could be controlled.

For a certain length of screw within the guide the standing wave indicator showed that the reflected wave had disappeared.

It follows that, if $\boxed{y_1}$ is the normalised admittance of the screw itself then, since,

$$\boxed{y} = (1 - j.1.07) \div \boxed{y_1} = 1$$

$$\boxed{y_1} = +j.1.07.$$

This length of screw therefore possesses a capacitive susceptance $\boxed{B_1} = 1.07$. To check the value of this susceptance the end of the waveguide was closed by a metal plate and the screw removed. A standing wave indicator located an E antinode and the screw was placed at this position projecting to the same extent into the guide as before.

The standing wave pattern was found to be displaced as a whole a distance 1.8 cm. away from the generator by the introduction of the screw; that is, the E antinode lay 1.8 cm away from the screw on the side away from the generator.

If we form $n' = 1.8/\lambda_g = 0.13$ and locate the point P (Fig. 234 (a)) on the real axis at the end of the n-arc whose n value is $n = n' = 0.13$ we obtain the normalised susceptance of the screw as $OP = j.1.07$. This follows because a displacement of $(0.5 - 0.13)\lambda_g$ from the screw towards the generator leads to a position of zero susceptance (E antinode). The corresponding motion on the circle diagram is along the imaginary axis from P upwards to plus infinity and then from minus infinity to 0.

It is found that when the screw is short its susceptance is positive (capacitive) but when it extends across the guide almost to the opposite face its susceptance becomes negative (inductive). There is an intermediate resonant length (approximately $\lambda/4$) at which the susceptance is zero. A wire extending across the centre of the waveguide with its length parallel to E behaves as an inductance in shunt.

43. Matching Devices

Various matching devices are employed in waveguides, most of them being of the symmetrical type which in effect introduces a susceptance in parallel with the guide. The only series obstacles in common use are slots in the walls, used for purposes of radiation and switching and these are dealt with elsewhere (Aerials, Chap. 17 Sec. 54).

Although in general, the only method of finding the susceptance of an obstacle is to measure it, yet the principles of electromagnetism have been applied to calculate the susceptances of several obstacles possessing simple geometrical forms. This makes it possible to design a structure to have a specified susceptance. A common form of obstacle is the so-called Iris which is formed by a thin metal strip or a pair of strips lying in the cross section of the waveguide and stretching from one wall to its opposite.

The principal types are :-

- (i) Capacitive Iris (Fig. 235)
- (ii) Inductive Iris (Fig. 236)
- (iii) Resonant Iris (Fig. 241)

The Capacitive Iris (H_{01} -mode)

In this case the strips are normal to the electric field as shown in Fig. 235. The formula for the normalised susceptance of this Iris is :-

$$[Y_1] = j[B_1] = j \frac{4a}{\lambda_g} \log_e \operatorname{cosec} \left(\frac{\pi d}{2a} \right).$$

The field in the vicinity of the edges of the strips resembles the electrostatic field that would exist were the side walls of the guide to be removed and the upper and lower faces to become a parallel plate condenser.

This quasi-electrostatic storage field can be represented mathematically as a set of evanescent E-waves.

However, under the influence of the incident

H_{01} -wave, oscillatory charging currents flow into and out of the strips from the upper and lower walls of the waveguide and these currents radiate the scattered H_{01} -waves. The presence of the storage field

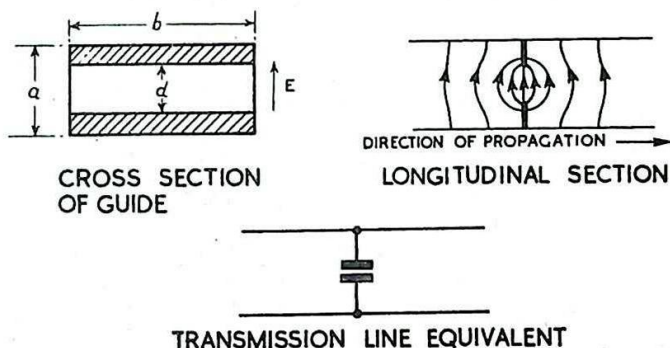


Fig. 235 - Capacitive iris (H_{01} -wave).

introduces a phase difference between these oscillatory currents and the oscillations of the incident E-field, so that a phase shift ϕ is produced in the reflected wave relative to the incident wave.

The capacitive iris is not employed where high powers are to be handled since the concentration of electric field at the edge tends to cause breakdown in the dielectric (air).

The Inductive Iris (H_{01} -mode)

Here the strips run parallel to the electric field, as illustrated in Fig. 236.

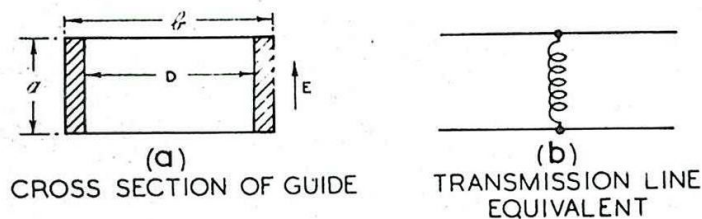


Fig. 236 - Inductive iris (H_{01} -wave).

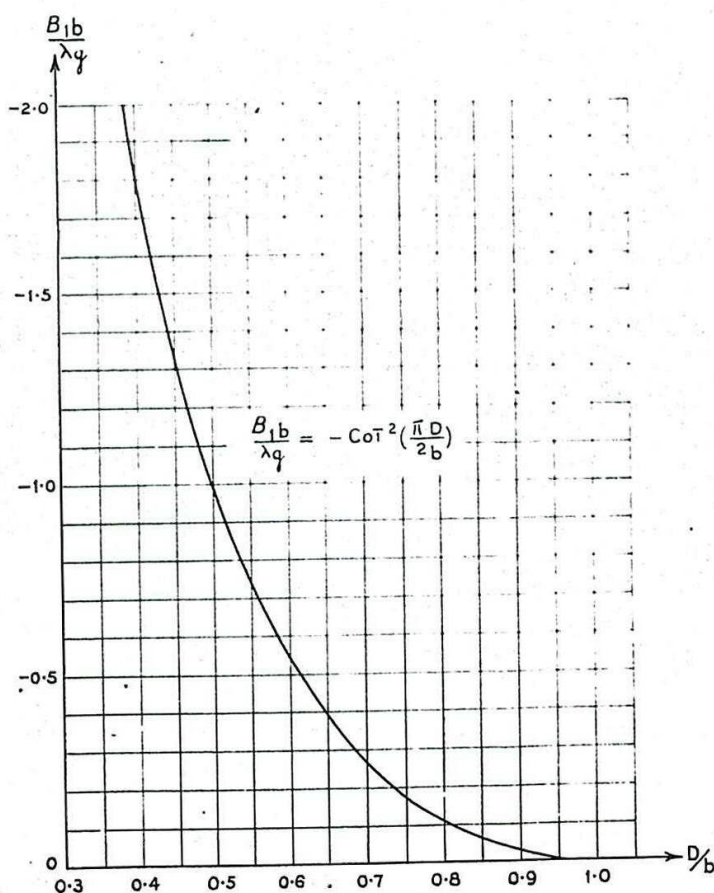


Fig. 237 - Design curve for inductive iris in rectangular guide (H_{01} -wave).

The formula for the susceptance of this iris is

$$Y_1 = -j B_1 = -j \frac{\lambda_g}{b} \cot^2 \left(\frac{\pi D}{2b} \right).$$

Fig. 237 is a design curve in which $\frac{B_1 b}{\lambda_g}$ is plotted as a function of D/b . The currents driven in the strips by the incident E vector excite a storage field resolvable into evanescent H-modes and at the same time radiate an H_{01} -wave, part of which is reflected back along the guide. This iris behaves like an inductance shunted across a transmission line. It is used in practice to eliminate the wave reflected from the open end of a waveguide.

Inductive Wire (H_{01} -mode)

A cylindrical wire stretched across the waveguide parallel to the E field behaves as an inductive shunting susceptance. The susceptance of the wire shown in Fig. 238 is, when the wire is thin ($r \ll \lambda_g$)

$$Y_1 = -j B_1 = -j \frac{2\lambda_g}{b} \frac{1}{\left(\log_{\epsilon} \frac{2b}{\pi r} - 2 \right)}.$$

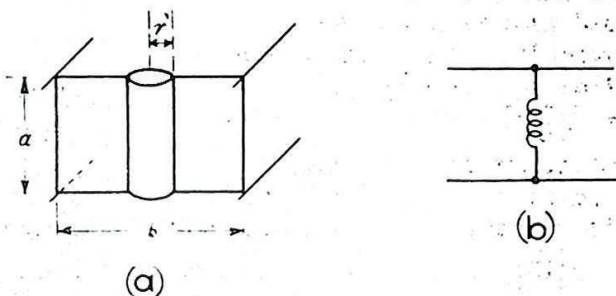


Fig. 238 - Inductive wire (H_{01} -wave).

Irises in Circular Waveguides (H_{11} -mode)

When a circular waveguide is designed to carry an H_{11} -wave but no other, then suitable diaphragms can be used to eliminate reflections as in rectangular waveguides.

Figs. 239 (a) and 239 (b) illustrate respectively the forms assumed by a capacitive and an inductive iris in a circular waveguide carrying an H_{11} -wave. The shaded regions represent thin metal diaphragms that lie in the plane of the cross-section.

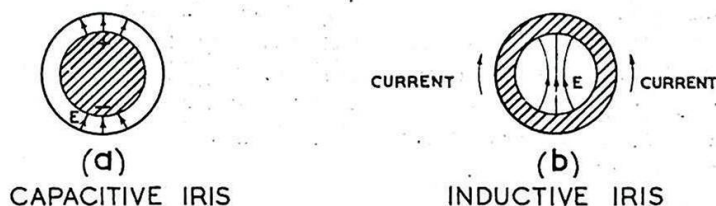


Fig. 239 - Irises in circular guides (H_{11} -wave).

Resonant Irises

We next consider how it is possible to combine a capacitive and an inductive iris to produce a composite iris whose admittance is zero and which therefore does not reflect the wave. Consider first the arrangement of irises indicated in Fig. 240, for an H_{01} -wave in a rectangular guide.

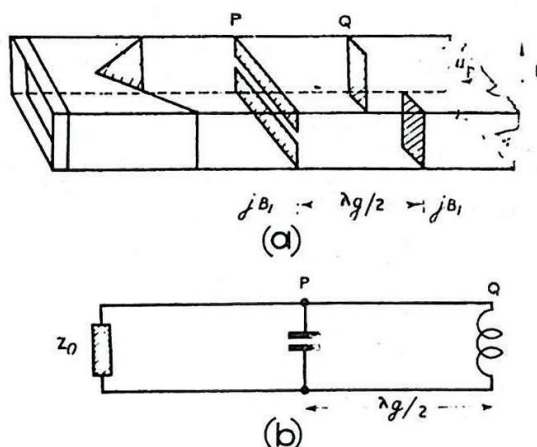


Fig. 240 - Resonant irises (H_{01} -wave).

The waveguide, which ends in a reflectionless load, carries a capacitive iris at the section P. The admittance of this section is $(1 + j[B_1])$ where $[B_1]$ is the susceptance of the iris. The admittance of the waveguide section at Q a distance $\lambda_g/2$ nearer the source than P is also $(1 + j[B_1])$ before the introduction of the inductive iris. When an inductive iris whose susceptance is $-j[B_1]$ is introduced at B the admittance of the section at Q becomes

$$[Y] = 1 + j[B_1] - j[B_1] = 1$$

Its reflexion coefficient is zero and no reflected wave returns to the source. In fact the iris at B has been used to match the waveguide at this point. Fig. 240 (b) shows the equivalent transmission line system.

There is clearly an infinity of pairs of irises ($\pm j[B_1]$) that can be combined in this way to produce a reflectionless combination. It would clearly be more convenient however if the two irises could be combined at the same section, Q for instance, instead of at separate sections P and Q. It is, in fact, found possible to superimpose a capacitive and an inductive iris to produce a reflectionless rectangular structure such as that shown in Fig. 241 (a). Such is known as a Resonant Iris. Its admittance is zero. It is found that the widths of the capacitive and inductive portions of the resonant iris are not the same as those required to give no reflection with the arrangement of Fig. 240 (a). This is because the storage fields of the two irises are now intermingled and the inductive and capacitive portions do not behave independently of each other. A useful, but approximate, design procedure is illustrated in Fig. 241 (b), where PQRS represents the section of the rectangular waveguide. The curves PLS and QMR are the two branches of a hyperbola that passes

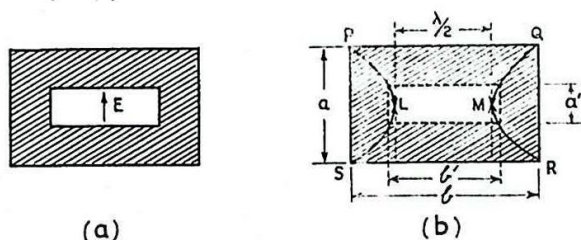


Fig. 241 - Resonant iris (H_{01} -wave).

through the corners PQRS and such that the distance between the poles IM is $\lambda/2$ (λ = free-space wavelength). If the corners of the aperture of the composite iris are made to fall on this hyperbola then the iris is approximately resonant. The dotted rectangle in Fig. 241 (b) is an example. There is an infinity of such resonant structures for any waveguide section.

If b and a , and b' and a' are the dimensions respectively of the waveguide cross-section and of the iris, then the geometrical construction given above is equivalent to the following relation between these dimensions.

$$\frac{b}{a} \sqrt{1 - \left(\frac{\lambda}{2b}\right)^2} = \frac{b'}{a'} \sqrt{1 - \left(\frac{\lambda}{2b'}\right)^2}$$

It may be seen from Fig. 241 (b) that when a'/b' is small, then b' is approximately equal to $\lambda/2$; that is, the resonant length of any narrow slot centrally placed in a diaphragm, with its length perpendicular to the electric field, is very nearly $\lambda/2$.

The transmission line equivalent of a resonant iris is shown in Fig. 242. At resonance the L-C circuit becomes a rejector circuit that places zero shunt admittance across the line. Consequently a progressive wave passes the structure without reflection. A thin $\lambda/2$ slot cut in a metal diaphragm across a circular guide carrying an H_{11} -wave and placed with its length at right angles to the electric field behaves as a resonant iris that transmits the incident wave completely.

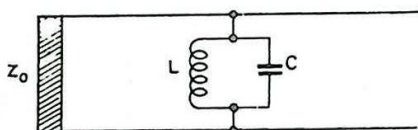


Fig. 242 - Equivalent circuit for resonant iris.

Fig. 243 - Transparent structures.

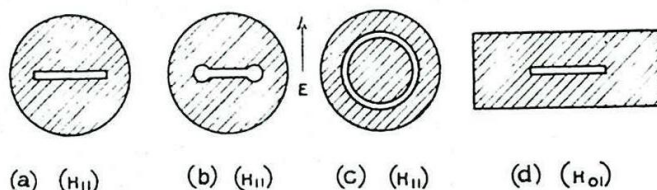


Fig. 243 shows a set of resonant structures that transmit the progressive wave without reflection. The resonant length of the slot in Fig. 243 (a) was found at $\lambda = 9.1$ cm. to be about 4% shorter than $\lambda/2$. For the composite L-C resonant iris of Fig. 243 (c) the inner circumference of the gap at resonance, when thin, is almost equal to λ .

The Q-Factor of Resonant Irises

Although the iris at resonance is almost perfectly transparent to the mode for which it is designed, changes in the frequency of the wave propagated along the guide cause changes in the admittance of the iris and partial reflection occurs. The behaviour is similar to that exhibited by a parallel resonant circuit shunted across a transmission line.

As shown in Chap. 1., Sec. 19 , the amplitude \hat{V} of the voltage developed across a parallel resonant circuit when supplied with a constant feed current is given by

$$\hat{V} = \hat{V}_r \cos \phi,$$

where $\tan \phi = 2 Q \delta$, δ being the fractional detuning, $\frac{f - f_r}{f_r}$ and \hat{V}_r the value of \hat{V} at resonance. It is thus possible to measure the Q of a parallel resonant circuit in terms of the reduction in the amplitude of \hat{V} for a given value of δ .

A similar procedure enables the Q of a resonant iris to be determined experimentally. The quantity measured is the magnitude of the voltage or current induced in a circuit coupled to the fields developed in the guide on the side of the iris farther from the generator, by means of a cable terminated in a non-resonant loop or probe. Thus, when Q is large the iris begins to give rise to serious reflections for a small fractional frequency shift $\pm f/f_r$, but if Q is small then it remains transparent over a reasonably large range of frequencies.

As example , consider the resonant structures shown in Fig. 243. At a wavelength of 9.1 cm. it was found that the Q of the slot of Fig. 243 (a) was 25 for a slot width of 0.5 mm but equal to 50 for a width of 0.1 mm. In general, the narrower the slot the larger is the value of Q .

The value of Q may be reduced by using the structure of Fig. 243(b) instead of a slot. With the straight section 4.2 mm. wide and the radius of the end circles equal to 1.4. mm the Q was reduced to the value 9. Such a structure therefore is transparent over a relatively wide range of frequencies.

The Q values in terms of ring width for the iris of Fig. 243 (c) were as follows :-

Ring width in mm.	0.1	0.5	0.8
Q	40	20	16

In all these examples the diaphragms were made of foil 0.004" thick. When brass of thickness $\frac{1}{8}$ " was used the Q factors were approximately doubled.

44. Switches and Protective Devices.

The electric field strength at the centre of a resonant slot is many times greater than that in the progressive waves at some distance from the slot. Consequently, when high powers are transmitted an electrical discharge may occur in the dielectric across the gap. This fact is made use of in the gas-filled resonant cell. Here the slot is enclosed in a glass capsule which is filled with commercial neon at a low pressure.

Fig. 244 (a) represents a sectional view of Fig. 243 (c), and Fig. 244 (b) a sectional view of Fig. 243 (a), each representing a resonant iris enclosed in a gas-filled cell. These cells with the diaphragms

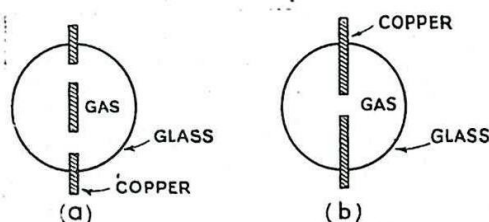


Fig. 244 - Simple T/R cell.

projecting as rims can be fitted into circular waveguides. They are quite transparent to waves of low power, such as the signal to the receiver, but spark over and become completely reflecting when a powerful wave from the transmitter strikes them.

A cell containing a resonant iris of the form shown in Fig. 243 (b) is used at 3 cm wavelengths under the name CV.115. In it, however, the circles are larger and the intervening straight portion of the window is much shorter than as shown in the figure.

Since cells are used to provide automatic switching of power in common T/R systems with waveguides, they are often referred to as T/R cells. A disadvantage of the simple cell of the type CV.115 is that when used for receiver protection the spark gap cannot be relied upon to strike immediately. When lag in striking occurs enough power may pass to the receiver to burn out the crystal.

A modified form of cell (American type 1B.24, British type CV.221) which affords satisfactory protection of the crystal is shown, diagrammatically, in Fig. 245. It is used at 3 cms wavelengths and breaks down at weaker field strengths than those required for the CV.115 and with no appreciable lag.

The cell is a gas-filled resonant cavity placed in series with the waveguide to which it is coupled at each end by slots. The cell cavity is separated from the two portions of the waveguide by glass windows. It is brought to resonance by adjusting the separation of the two spikes that project into it as shown, the upper spike being hollow. This adjustment is accomplished by means of a screw which works

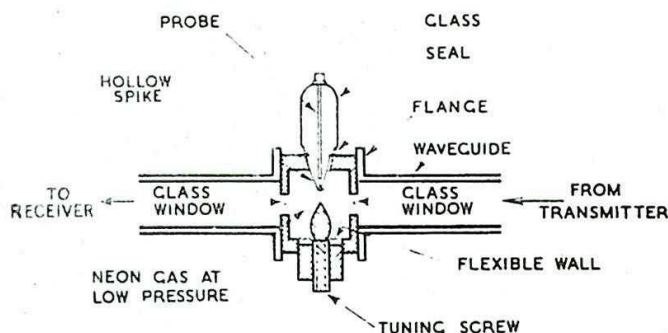


Fig. 245 - Adjustable T/R cell.

against a flexible area of the wall that carries one of the spikes. To avoid lag in the striking of the discharge when a power wave enters the cell, a glow discharge is kept running between the probe and the inner wall of the upper spike. This spike has a hole in the end and electrons diffuse from the glow discharge into the space between the spikes to provide the initial ionisation from which the high frequency discharge is able to build up without lag. A high resistance is included in series with the probe to limit the glow discharge, but its value must be chosen to avoid an intermittent discharge, for then the protective action might be vitiated; (compare Flashing Neon, Chap. 10 Sec. 1 and Chap. 19 Sec. 10). The cell permits weak signals to pass unaffected but blocks strong signals.

A soft rhumbatron may also be used as a T/R cell, in a manner similar to that employed in transmission line circuits. Instead of the resonant cavity being coupled by means of pick-up loops, as described in Chap. 4; windows are used, the whole cell being tuned by the cavity tuning adjustment. The arrangement is further described in Sec. 50.

45. Reflectors in Waveguides

Various reflecting irises are depicted in Fig. 246.

When each is compared with the corresponding resonant iris of Fig. 243, it will be observed that the reflectors are obtained by an interchange of the metal and the open portions of the diaphragms followed by a rotation through a right angle. A pair of diaphragms related in this way are said to be complementary. Not only are their geometrical properties complementary, but it is found both by theory and practice that their electromagnetic properties are complementary also. Thus,

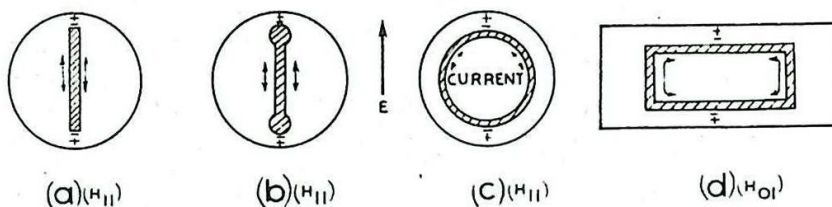


Fig. 246 - Reflecting irises.

the electrical behaviour of the thin ring in Fig. 246 (c) may be compared with that of its counterpart, the thin circular slot of Fig. 243 (c). When the circular slot has an inner circumference approximately equal to the wavelength λ it is transparent and has zero susceptance. It is found, on the other hand that when the ring of Fig. 246 (c) is made slightly greater than λ in circumference its susceptance becomes infinite and the section containing it becomes completely reflecting. Similarly, the halfwave strips of Fig. 246 (a) and (b) are reflectors to an H_{11} -wave.

Reflecting rings form very convenient mechanical switches for diverting power alternately from one branch in a waveguide to another. The small inertia of the switches allows them to be turned at high speed. Two examples of the use of ring switches are given in Fig. 247. The reflectors must be positioned as shown in Fig. 246 in order to reflect almost completely the incident wave and must be turned through a

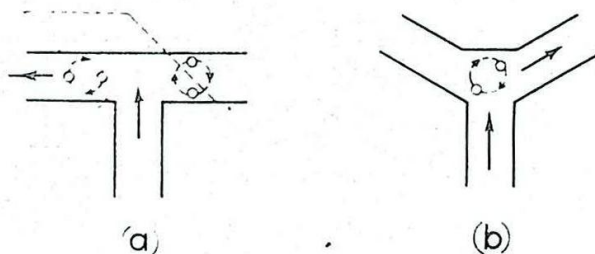


Fig. 247 - Use of rotating iris for switching.

right angle so that they are perpendicular to the electric field if they are to allow the incident wave to be propagated undisturbed. This may be achieved if the iris is rotated about the horizontal transverse axis. The mechanical simplicity of this arrangement makes it particularly suitable for use as a reflecting switch. Sometimes in rectangular waveguides a reflecting switch is made in the form of a rectangle (Fig. 246 (d)) to fit into the waveguide cross section. Its total perimeter is still of the order of one wavelength.

REFLECTIONS FROM JOINTS AND BENDS IN WAVEGUIDES

46. General

In radar a magnetron oscillator is often connected directly to a waveguide so that the circuit load on the magnetron is the impedance which the waveguide presents at its input end. It is explained in Chap. 8 Sec. 41 that the frequency of oscillation of a magnetron is peculiarly sensitive to the nature of the load. If therefore any irregularity is present in a waveguide, so as to cause a reflected wave, it may produce an effective impedance at the magnetron end which will cause the oscillation to take place at some undesirable frequency. This control of the frequency of oscillation by the attached waveguide is called Frequency Pulling. When a long run of waveguide is employed an additional effect called Frequency Splitting is observed when the termination reflects a wave down the guide. This effect does not appear when l/λ_g is small.

For these reasons it is desirable to keep all waveguide runs short and to design all joints and bends so that they cause the minimum of reflection.

47. Waveguide Joints

Flanges. At wavelengths of 9 and 10 centimetres coupling between two sections of waveguide is achieved through flat end-flanges, as shown in Fig. 248 (a) and (b).

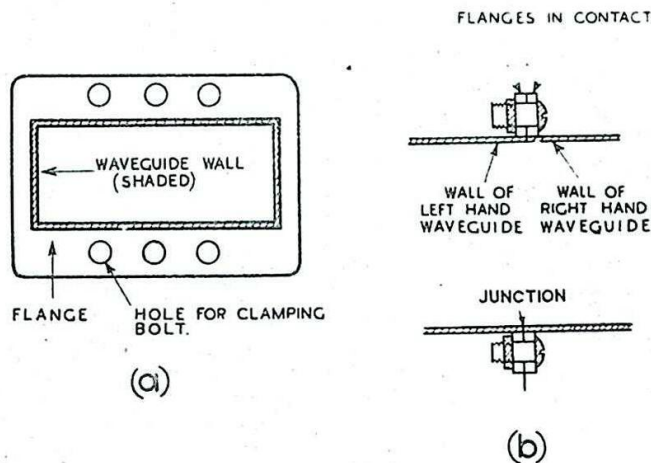


Fig. 248 - Flanges for connecting waveguide sections.

Fig. 248 (a) shows the face of the flange fitting flush with the end of the guide and 248 (b) two waveguides clamped together by their flanges.

The principal causes of reflection at these junctions are :-

(i) Misalignment of the walls at the junction causing a step in the walls of the guide. To avoid this the bolt holes must be located accurately.

(ii) Gaps between the waveguide walls across the junction either due to imperfections in the plane surfaces of the flanges or

because they are not flush with the ends of the guides. Fig. 249 indicates what may occur when the flange faces are not accurately plane. The flanges are in contact at C but have left a gap at G between the waveguide walls. The two flange faces between G and C may be regarded as a pair of transmission lines of length ℓ and short-circuited at C, so that if by chance ℓ is approximately equal to $\lambda/4$ (λ = free-space wavelength) an E antinode and an H-node occur at the gap G. There is therefore a discontinuity in the transverse component of H in the H_{01} -wave in

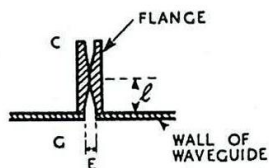


Fig. 249 - Irregularity due to flanges.

the guide at G, and therefore also in the longitudinal wall current. Such a gap causes a serious reflection. Consequently, the faces of the flanges should be made as nearly as possible accurately plane and parallel. Alternatively a copper gasket fitted between the flanges may be used to ensure good contact.

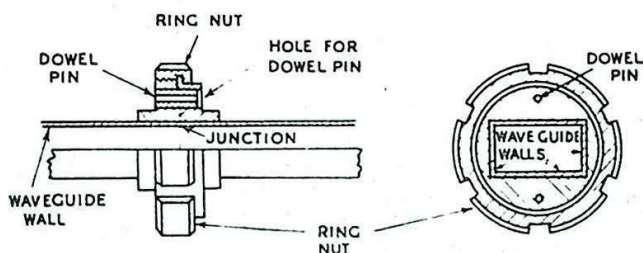


Fig. 250 - Coupling unit for $\lambda = 3$ cms.

At a wavelength of 3 centimetres these effects (i) and (ii) are more difficult to avoid than at the longer wavelengths and the design of the coupling unit is somewhat more complicated than that described above. A suitable device is shown in Fig. 250. The two flanges shown here are of more solid construction and are clamped together by a ring nut which presses on the one (female) and screws on to the other (male). A step at the junction is avoided by accurately placed locating pins.

As mentioned above and illustrated in Fig. 249 a common source of reflection at a joint is the existence of a gap between the walls of the two lengths of waveguide. A form of coupling which is finding increasing favour uses the electrical properties of a space between the flanges to provide a theoretically perfect electrical union between two lengths of waveguide.

The principle of these double quarter-wavelength joints, commonly called choke or capacity joints, is shown in Fig. 251. An L-shaped recess GBA is formed by suitably shaping the surface of one of the flanges, the other remaining plain. This recess follows the contour of the section of the waveguide wall and may be employed with either circular or rectangular waveguides.

The lengths BG and BA are each $\lambda/4$ (λ is the free-space wave-length) as indicated. The portions of the recess GB and BA each form quarter-wavelengths of transmission line; consequently, the short-circuit at A is transferred to G where the standing wave in the recess produces an H-antinode. The transverse component of H in the wave thus remains continuous across the gap. Hence, equal longitudinal currents flow into and out from the gap as indicated, and no reflected wave is generated in the waveguide at G.

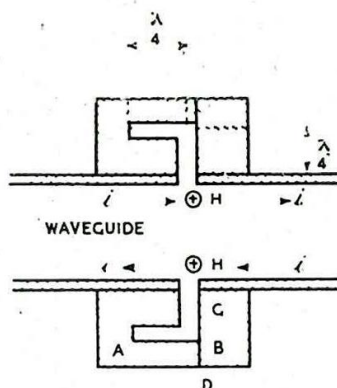


Fig. 251 - Double $\lambda/4$ coupling unit,

The same principle is used to approach the ideal of a perfectly reflecting piston or plunger in circular and rectangular waveguides. It is difficult to ensure, with the normal piston, that good surface contact exists around the whole perimeter without at the same time rendering the piston stiff in action.

The Double $\frac{\lambda}{4}$ Tuning Plunger indicated in Fig. 252 achieves both easy movement and good electrical termination. The recesses AB and BC are both $\lambda/4$ lengths of transmission line systems, and as before ensure that across the gaps between the face of the piston and the waveguide wall (sections at A and D) there is an antinode of H and no component of E tangential to the face. The whole face, including the gap, behaves therefore as a perfectly fitting and reflecting disc which reflects the wave in the guide without generating evanescent modes.

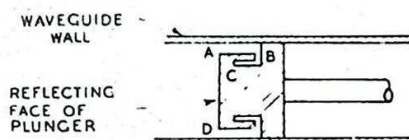


Fig. 252 - Double $\lambda/4$ tuning unit.

48. Bends in Waveguides

In practice it is necessary to introduce corners and bends into waveguide runs by the insertion of sections with the forms illustrated in Fig. 253. The bend shown at (a) is known as an H-bend and that of (b) an E-bend. Fortunately, it is found that provided the inner radius of the bend exceeds λ_g , and the section of the guide is not distorted in the process of bending, the reflection produced as an H_{01} -wave enters is very small. The standing wave ratio associated with a good bend is of the order of 1.05.

The wave in the bend resolves into two waves, and in the case of waveguides with a circular section carrying an H_{11} -wave these waves travel in the bend at different speeds; consequently when they emerge to recombine into an H_{11} -wave, the plane of polarisation of the issuing wave is rotated with respect to that entering. This is a disadvantage of circular waveguides.

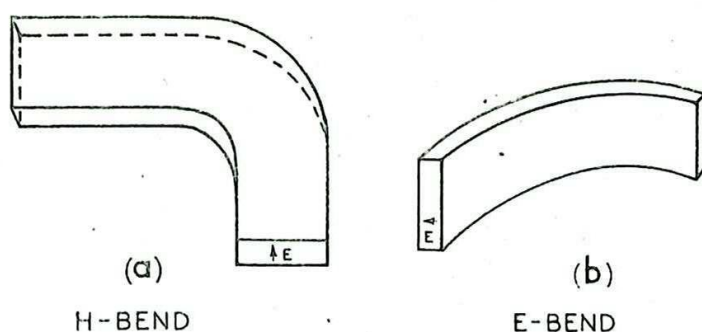


Fig. 253 - Smooth bends in waveguides.

When it is necessary to economise in space, the smooth bends discussed above are sometimes found to be inconvenient, especially at the longer wavelengths; consequently the sharp corners depicted in Fig. 254 are commonly employed. Fig. 254 (a) depicts an H-corner and (b) an E-corner, and it is to be noted that in each case the outer corner of the bend has been sliced off and the hole formed in the walls closed by a flat plate. Experiment shows that for each bend there is an optimum value of the ratio c/d that gives negligible reflection. This value for an H-corner was found to be

$$c/d = 0.65,$$

and for an

E-corner $c/d = 0.61,$

with a wavelength of 10.8 cm. in a 7 cm. x 3.25 cm. waveguide.

Twists A rectangular waveguide may be twisted so that the plane of the E-vector is rotated through an angle (usually 90°), and little reflection arises provided the length of twist is not small compared with λ_g . Such a twist is shown in Fig. 255.

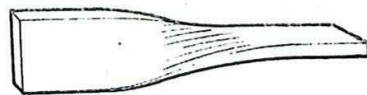


Fig. 255 - Twist in waveguide.

49. Rotary E-H Transformer

The aerial systems of most centimetre wave equipments comprise a reflector which is a

portion of paraboloid or parabolic cylinder, and is capable of motion about horizontal and/or vertical axes. Power is fed to the aerials from the end of a waveguide which in some instances is terminated in a horn. The waveguide feed must move with the mirror but must also be coupled to the fixed waveguide run from the transmitter. The aerial feeder must be coupled to the main waveguide so that no variation in power or polarisation of the radiated pulse occurs as the mirror moves. This problem is solved by use of special rotary transformers. These transformers take advantage of the fact that an E_{01} -wave in a circular guide is readily excited by an H_{01} -wave in a rectangular guide which feeds into the wall of the circular guide so that the E vector in the H_{01} -wave is parallel to the axis of the circular guide.

Fig. 256 illustrates the principle of this device. At (a) we have an H_{01} -wave in a rectangular guide, running into the lower end of a closed circular guide whose radius is large enough for the guide to accept a progressive E_{01} -wave. It is arranged that the electric vector in the H_{01} -wave is parallel to the axis of the circular tube and readily excites in it the E_{01} -mode as a progressive wave which travels to the top of the circular guide as shown. The longitudinal E-vector in the circular tube excites an H_{01} -wave in the second rectangular guide leading to the aerial.

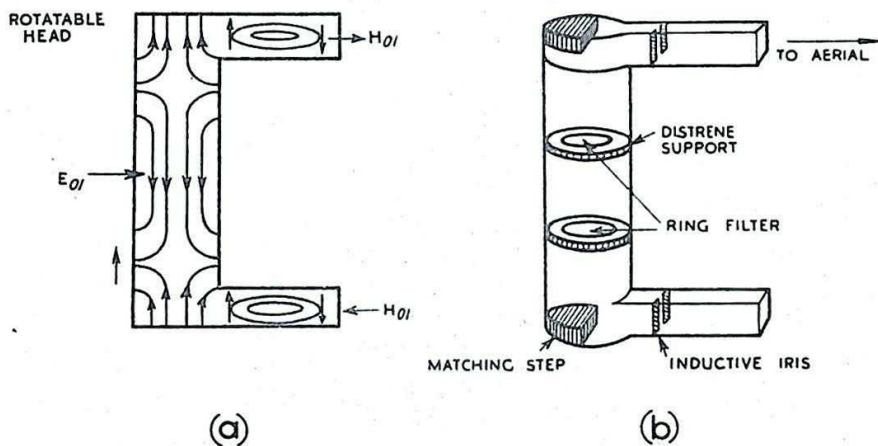


Fig. 256 - Rotary E-H transformer.

At (b) certain ancillary devices are also shown; they are, a pair of ring filters to reflect the H_{11} -mode which can also travel in the circular tube, and matching steps and inductive irises to eliminate reflections at the junctions of the circular guide and the rectangular guides, both in transmission and in reception.

If the top of the circular tube is made rotatable then the rectangular guide can be

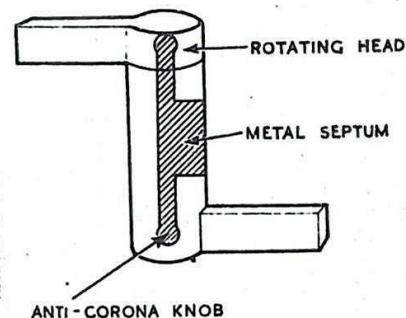


Fig. 257 - Rotary septate-coaxial transformer.

swung in azimuth. Since the E_{01} -wave possesses axial symmetry it will feed equally well into the rectangular guide at all azimuths.

An alternative rotary joint employs a septate-coaxial combination which permits a considerable reduction in the overall diameter of the circular tube (Sec. 21). A diagrammatic representation of this form of rotary joint is given in Fig. 257.

50. COMMON T/R WITH WAVEGUIDES

In Chap. 4 Common T/R circuits are described, applicable to transmission lines. In radar systems employing waveguides, the corresponding waveguide connections must be introduced in order to provide common T/R working. To reduce the waveguide problem to the same basis as that of the transmission line, consider the currents, at any instant, in the walls of the guide (Fig. 207), which is taken to be rectangular and carrying an H_{01} -wave.

Referring to the figure there are two current flows to be distinguished.

Some current-flow lines are entirely confined to the wide walls of the guide and are directed, generally speaking, in the direction of propagation. These currents are strong along the centre line of the wide sides and weaken considerably for small deviations from the centre. Other current-flow lines start near the centre of a wide side and trace out a path over the narrow side and on to the wide side opposite to that from which they started. These current-flow lines are, generally speaking, transverse to the direction of propagation of power down the guide. The first type of current line corresponds to the current in a twin balanced transmission line,

which may in this case, be taken as a pair of parallel ribbons. The second, transverse, current line corresponds to the current in a shunt short-circuited stub across a twin transmission line. The circuit equivalent of a rectangular waveguide carrying an H_{01} -wave is, therefore, a twin transmission line modified by the addition of a large number of shunt stubs. This is shown in Fig. 258. If the stubs are not to upset the transmission, they must be $\lambda/4$ long.

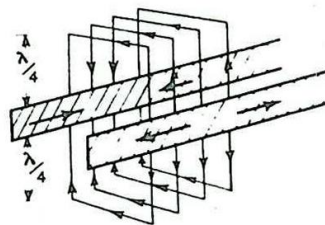


Fig. 258 - Transmission line equivalent of waveguide walls.

It is now possible by analogy to work out the common T/R connections. A soft rhumbatron (Chap. 6) is employed as a switch. This acts like a parallel tuned circuit of fairly high Q and will spark over readily. In Chap. 4 (Fig. 182 (a)) the rhumbatron is shown with loop coupling, but for coupling to waveguides windows or irises are used. A hole is cut in the side of the rhumbatron and a similar hole in the guide wall; the rhumbatron being then clamped firmly against the guide (Fig. 259). With a suitable size of hole the coupling is practically 1:1 and the rhumbatron; or parallel tuned circuit, may then be considered as directly connected. Thus if a rhumbatron, with window coupling, is placed on the narrow side of a rectangular guide (Fig. 259), the equivalent circuit is as shown in Fig. 260, the parallel tuned circuit and switch being $\lambda/4$ away from the transmission line and shunt connected. The shunt arrangements of Figs. 180, 181 and 182 (Chap. 4) can thus be developed.

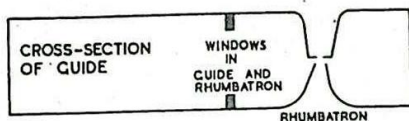


Fig. 259 - Shunt window coupling between rhumbatron and guide.

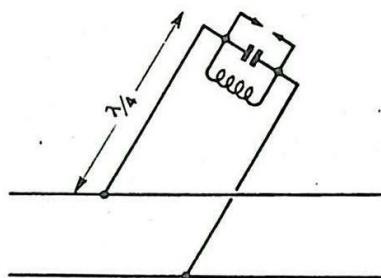


Fig. 260 - Equivalent circuit for rhumbatron switch.

A typical T/R connection with rectangular waveguide is shown in Fig. 261. Two rhumatrons are employed with window coupling to the guide. The receiver feeder is loop-coupled into the rhumbatron nearest the aerial, and the equivalent circuit is shown in Fig. 262. During transmission, the gas in the rhumatrons ionizes. The tuned circuits are thus short-circuited, giving rise to open circuits $\lambda_g/4$ away on the main line

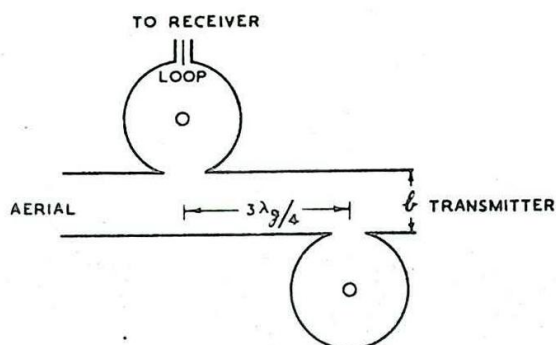


Fig. 261 - Side view of common T/R switch, employing two soft rhumatrons shunt-connected.

and allowing the transmitter power to pass to the aerial. During reception the gases are deionised. The rhumbatron nearer the transmitter acts like an open circuit and presents zero impedance $\lambda/4$ away, on the main line. This apparent short-circuit seen from the receiver junction $3\lambda_g/4$ away, looks like an open-circuit. The received signals therefore find a very high impedance, looking towards the transmitter, at the receiver junction, and proceed through the receiver rhumbatron and into the receiver feeder.

On very short wavelengths, the feeder connection from the rhumbatron to the receiver may be undesirable owing to excessive attenuation in the feeder. In this case a waveguide replaces the feeder and there is window coupling both into and out of the rhumbatron. Series connection to the waveguide may also be used instead of shunt. In this case the rhumbatron is connected to the broad face of the guide and is usually off-set effectively $\lambda/2$ from this face by interposing another section of waveguide (Fig. 263). This section and the guide leading to the receiver are often circular. They carry an H_{11} -wave and so are similar to a rectangular guide carrying an H_{01} -wave. The

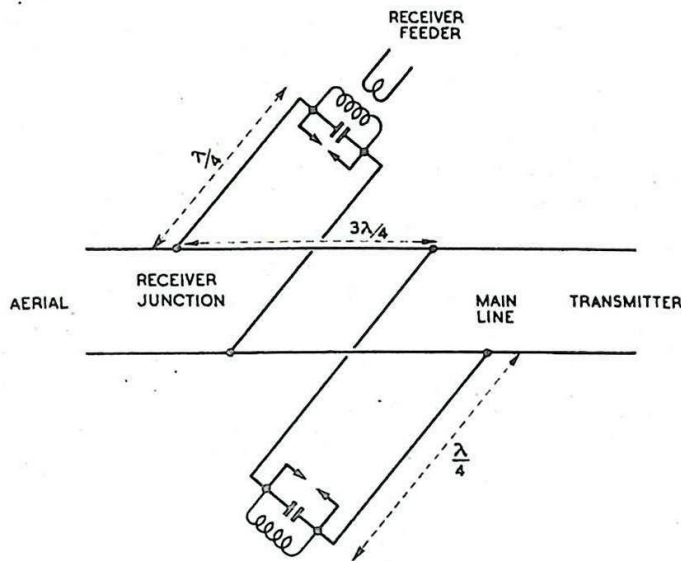


Fig. 262 - Equivalent circuit for
fig. 261.

circular guide is wider than the rectangular and so it is easier to place a crystal across it, to form the crystal converter which is the initial part of the receiver.

It is usually found in practice that the tuning of the rhumbatron nearer the transmitter is very broad and hardly affects the performance of the set. On very short wavelengths this rhumbatron is sometimes replaced by a resonant iris of the dumbbell type (Fig. 243 (b)). This has a low Q and would not be suitable as a replacement for the receiver rhumbatron.

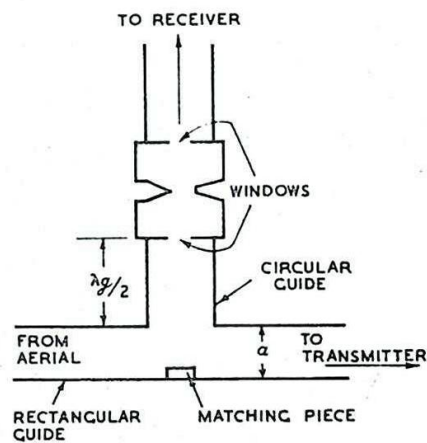


Fig. 263 - Series connection
of soft rhumbatron switch.

CHAPTER 6

ELECTRONIC DEVICES

CATHODE RAY TUBES

1. Construction and Operation

The Cathode Ray Tube (CRT) is the principal display device in most radar systems. It can be thought of as an indicating device with a very light pointer which has no inertia. The pointer is, in fact, a beam of electrons which can be deviated by means of electric or magnetic fields. Where the electron beam impinges upon the fluorescent screen of the tube it forms a spot of light, and as the beam is moved the spot traces out a pattern on the screen which is called the Trace. The brightness of the spot can be altered by the application of suitable potentials so that the trace may be made brighter in some parts than in others. In some tubes the trace on the screen is transitory and disappears as soon as the electron beam is displaced, while in others there is an afterglow period so that the trace remains visible for some time after the electron beam is moved.

It is instructive to compare the internal construction of a CRT with that of a triode. In a triode the control grid is used to vary the quantity of electrons reaching the anode so that the electron stream from the cathode is intensity modulated. In a similar way a cathode ray tube has a cathode to produce a stream of electrons which are accelerated by an anode system and can be intensity modulated by the variation of potential difference between the control electrode and the cathode. The electron stream, however, needs to be concentrated (focused) in a narrow beam so that after it has passed through the anode system it produces a small sharply defined spot of light.

The beam of electrons can be changed in direction by an electric or a magnetic field or can be modulated in intensity, or both, and the effects will be made visible on the fluorescent screen. If the beam of electrons is deflected rapidly and repeatedly the moving spot of light will produce a persistent trace on the screen. Intensity modulation of the beam will vary the number of electrons reaching the screen in a given interval of time and so alter the brightness (brilliance) of the spot or trace (Fig. 264).

The cathode ray tube consists essentially of:-

- (i) The electron gun (or, simply, the gun) comprising:- a cathode which acts as a source of electrons; a control electrode which alters the electron concentration in the beam; and an anode to accelerate the electrons.

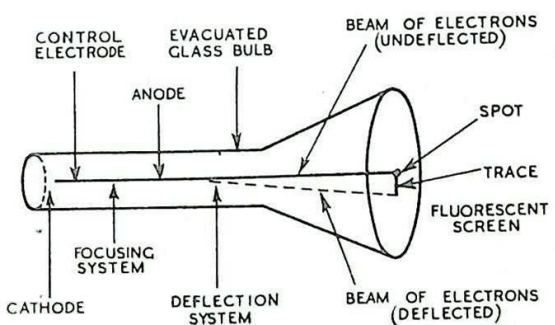


Fig. 264 - Schematic layout of CRT

- (ii) A focusing system which is commonly (in electrostatic focusing) part of the gun.
- (iii) A deflecting system.
- (iv) A fluorescent screen.
- (v) An evacuated glass bulb.

2. Types of Tubes

(i) The most common type of tube used in radar is the hard tube in which the glass envelope containing the electrode system is sealed off after the air pressure has been reduced to about 10^{-6} mm. of mercury. The accelerating potential may have a value from about 500 V to about 5 kV. depending on the size of the tube; tubes normally have screen diameters between 1 and 15 inches.

(ii) Soft tubes, in which the gas pressure is about 0.01 mm. of mercury, are also in use. They usually employ an accelerating potential whose value lies between 200 and 1000 V. This type of tube has serious disadvantages compared with the hard tube (Sec. 14) and is not commonly used.

3. The Cathode

The cathode is usually of the indirectly heated type and consists of a small nickel tube closed at one end. The emissive substance (such as a mixture of barium and strontium oxides) is situated in a depression in the closed end of the tube and the heater, supplied with an alternating voltage (from 4 to 6.3 V. depending on the particular CRT) is inserted through the open end of the tube. The beam current is of the order of 30 - 300 microamps. Directly heated filaments are not commonly used as cathodes except in soft tubes, since if they are supplied with an alternating voltage the electron stream is likely to be adversely affected.

4. The Anode

The anode is mounted a short distance from the cathode on the side nearer the fluorescent screen, and usually (when the anode is also part of the focusing system) takes the form of a disc with a hole about 1 mm. in diameter in the centre. When the anode is simply an accelerating electrode, it often takes the form of a conducting layer of carbon particles (aquadag coating) on the inside surface of the glass envelope of the CRT.

The potential of the anode is maintained at some hundreds or thousands of volts positive with respect to the cathode in order to accelerate the electrons. Some electrons travelling at a high speed from the cathode are collected by the anode, whilst some pass through the hole and continue to travel down the tube until they strike the fluorescent screen. The electron beam is divergent, due partly to the mutual repulsion of the electrons as they pass down the tube, and partly to the angle at which electrons from different parts of the emitting surface pass through the anode aperture. This divergence tends to cause the spot of light on the fluorescent screen to be large and diffuse. To make the spot small and well defined the electrons must be made to converge so as to strike the screen over as small an area as possible. A focusing system (Secs. 11 - 14) is used for this purpose.

5. The Control Electrode (Control Grid)

Only a small proportion of the total number of emitted electrons pass through the hole in the anode unless the electron stream from the cathode can be compressed into a narrow beam which can pass through the anode aperture. This compression is performed by the Control Electrode which usually takes the form of a hollow nickel cylinder surrounding the cathode. It is called the Shield, Wehnelt Cylinder or Grid, the last name being generally used because the electrode has a control on the beam current similar to that exerted by the control grid of a triode valve on the anode current. If a potential, negative compared with that of the cathode, is applied to the grid a region of minimum potential is produced in the neighbourhood of the cathode. The immediate effect of this potential minimum is to reduce the divergence of the electron stream. Increasing values of negative potential on the grid cause the electron stream to become more and more compact until at some optimum value the majority of the electrons pass through the anode aperture. However, if the negative potential is increased beyond this value there is an appreciable reduction in the beam current. In a valve the anode current is reduced in a similar manner. The brightness or brilliance of the spot on the CRT screen depends on the number of electrons reaching the screen in a given interval of time and so is controllable by the potential on the grid. The grid potential can be made sufficiently negative so that the grid field neutralises that of the anode. Then the beam of electrons is cut off and the spot of light on the screen is "blacked out". This corresponds to cut-off bias on the grid of a triode.

6. Brilliance

The intrinsic brilliance of the spot depends on the energy contained in the beam. It is, therefore, proportional to both:-

- (i) The square of the velocity of the electrons, which depends on the potential difference between anode and cathode, and
- (ii) the current density (number of electrons striking the screen per unit interval of time) of the beam which largely depends (Sec. 5) on the potential difference between the grid and cathode.

If the spot is swept over the screen the energy of the illumination is spread over a large area. Hence, on the assumption that the frequency with which the movement of the spot is repeated remains constant, the greater the magnitude or speed of this movement, the less the apparent brilliance of the trace. Thus, high speed and large tubes require high anode voltages in order to produce adequate brilliance. When the spot is stationary the brightness must be reduced to a minimum (by control of the grid-cathode voltage), otherwise the persistent electron bombardment causes a portion of the screen to lose its fluorescence; the tube screen is then said to be "burnt" and no fluorescence appears afterwards at this place.

7. The Fluorescent Screen

The screen consists of a translucent layer of fine powder adhering to the end of the tube. The emission of light during the actual stimulus of the beam is termed fluorescence; light which continues to be emitted after the stimulus has been removed is due to phosphorescence and is called Afterglow. The material of the powder determines the colour of the trace, and also the duration of afterglow.

The desirability or otherwise of using a screen with long afterglow depends on the purpose for which the tube is to be used.

The behaviour of the screen, including its luminous efficiency, depends greatly on the purity of the compound used. Such substances as calcium tungstate (blue-violet trace), zinc silicate (green), zinc phosphate (red) and preparations of zinc sulphide and zinc cadmium sulphide are a few from amongst those that may be used. A minute trace of an Activator such as silver or copper is necessary to produce the maximum luminous efficiency, which is of the order of 1 - 5 candle power per watt. In a good many cases afterglow of a screen is limited to a few microseconds, and by the addition of a suitable "killer" such as a compound of nickel, it can be cut down to a fraction of a microsecond. When it is necessary to examine a trace long after the stimulus giving rise to it has ended, a screen with a long afterglow of several seconds may be used; such a screen may consist of zinc sulphide with a copper compound as an activator.

8. Screens with Multiple Layers

In general it is found that when a screen has an afterglow of short duration the build-up time for the light to reach maximum intensity is short and the average intensity is high. However, when the screen has an afterglow of long duration the build-up time of the intensity of the light is comparatively long and the average intensity is low. In order to obtain long-afterglow characteristics with a higher average intensity of light it is possible to use a screen consisting of two layers of different fluorescent materials. The electron beam strikes the first layer and causes it to emit ultra-violet and visible blue light. The afterglow duration of this layer is usually a small fraction of a second. The ultra-violet light acts on the second layer (nearer the face of the tube) and causes the emission from it of visible light (usually yellow). The afterglow duration of this layer is of the order of a few seconds. An arrangement of this kind is found to give a comparatively high average intensity of illumination combined with long afterglow. Since it has been arranged in this case that the blue light is of short whilst the yellow light is of long duration, it is possible, with suitable light filters, to arrange for the screen to show the characteristics of either short or long afterglow. Thus if the movement of the spot is viewed through a filter capable of passing blue light only, then the short afterglow is seen. If, however a yellow filter is used, only the long afterglow characteristic is visible.

The use of screens with multiple layers allows for the possibility of distinguishing between those movements of the spot of light which are regularly repeated at a fixed position on the screen and those which are not. This possibility is of great importance in radar applications. Suppose for example that one layer of the screen, with short afterglow characteristics, emits red light when activated, whilst another layer, with long afterglow emits green light. If movements of the electron beam are regularly repeated, with-in a period shorter than the duration of the afterglow, and at a fixed position on the screen, the green light can grow in intensity in comparison with the red. If the movements of the beam of electrons are not repeated at the same position on the screen the green light at any one position will not attain any appreciable intensity and the briefer emission of red light will predominate.

9. The Skiatron or Dark-Trace Tube

Screens are also in use whose materials react to the stimulus of the electron beam by a change of colour rather than by fluorescing. This colouration is most marked in the case of certain of the alkali halides, and the final colour obtained varies with the salt used for

the material of the screen. In general the material used is potassium chloride, and the electron beam causes a dark magenta stain where it strikes the white screen.

The staining does not die away instantaneously on the removal of the exciting electron beam. There is in fact an "afterglow" - if such a term can be applied to a dark mark formed on a white background. The rate of the decay of the colour is found to depend mainly on:-

- (i) the initial intensity of the marking on the screen, the decay taking longer, the more intense is the initial mark;
- (ii) the intensity and colour of any light incident on the screen from an external source, the decay being more rapid the greater is the intensity;
- (iii) the temperature of the screen, a rise in temperature resulting in a more rapid decay.

If the face of the CRT is brightly illuminated, it and any dark trace on it can be projected as in an episcopo on to a large ground-glass screen. In this way magnified images of the stains on the face of the tube can be obtained. Mercury light is usually chosen as the illuminant for the screen for two reasons. In the first place, it is rich in the yellowish-green light band, a colour which is approximately complementary of the magenta stain, so that there is good contrast between the colour of the screen and that of the stain. Secondly, light of this colour is the most active in producing the decay of the stains. When the screen is illuminated with an intensity of about 70,000 foot-candles of mercury light the stains last for about 10 seconds.

The type of tube described above is commonly called a Skiatron.

10. The Glass Envelope

The end of the glass envelope, on which the fluorescent screen is deposited, has a curvature which is consistent with mechanical strength. It is worth noting that the end of a tube of diameter 12 inches carries a load of $\frac{3}{4}$ ton due to atmospheric pressure alone.

FOCUSING SYSTEMS

11. General

The necessity for focusing the electron beam was discussed in Sec. 4. There are three methods available:-

- (i) Electrostatic - with hard tubes.
- (ii) Magnetostatic - with hard tubes.
- (iii) Gas - with soft tubes.

Each of those three methods has its own advantages and disadvantages but (i) and (ii) are in common use in radar whereas (iii) is rarely used.

12. Electrostatic Focusing

This form of focusing utilises the effects of suitably shaped electric fields on the electron beam. The effects are very similar to those obtained when a beam of light passes through materials of

different optical refractive indices. In this connection a new branch of applied science, known as Electron Optics, has come into existence, and many of the terms used are borrowed direct from physical optics.

A ray of light is refracted if it passes from a medium of one refractive index to a medium of another. If the refractive index of the first medium is smaller than that of the second, the ray of light is bent towards the line normal to the boundary of the two media. If an electron, moving with velocity u_1 through a region of constant potential V_1 passes into a region of constant potential V_2 its velocity is altered and its path changes direction at the boundary between the two regions. If u_1 (proportional to the square-root of V_1) is less than u_2 , (proportional to the square-root of V_2) the path of the electron is bent towards the line normal to the boundary: (Fig. 265).

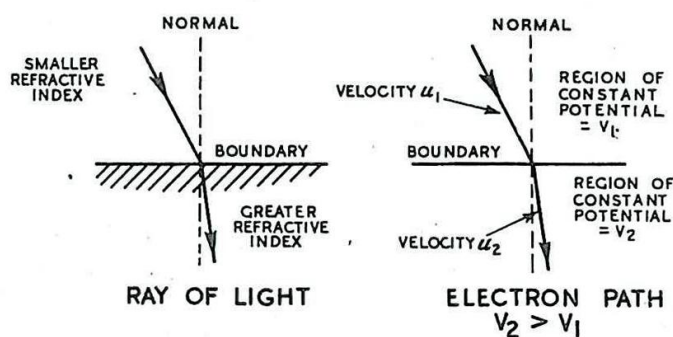


Fig. 265 - Refraction of ray of light and of electron beam.

In most cases considered in electron optics the change of potential from one region to another is gradual and the path of the electron is curved. In Fig. 266 the electron is shown passing through a region of constant potential gradient, the direction of its initial movement making an angle with the lines of electrostatic force. In electron optics the equipotential planes, all at right-angles to the lines of electrostatic force, are equivalent to surfaces of constant refractive index in physical optics. The electron undergoes continuous refraction as it passes through the equipotential surfaces, and if it is moving from regions of lower to regions of high potential, its path always tends to coincide more and more with a direction perpendicular to the equipotential planes. This last statement is true whatever the shape of the equipotential surfaces. In an accelerating field an electron is deflected towards the normal to the equipotential surfaces; in a retarding field it is deflected away from the normal.

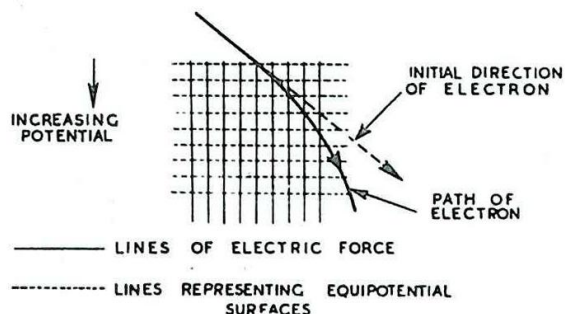


Fig. 266 - Path of electron moving through uniform accelerating electric field.

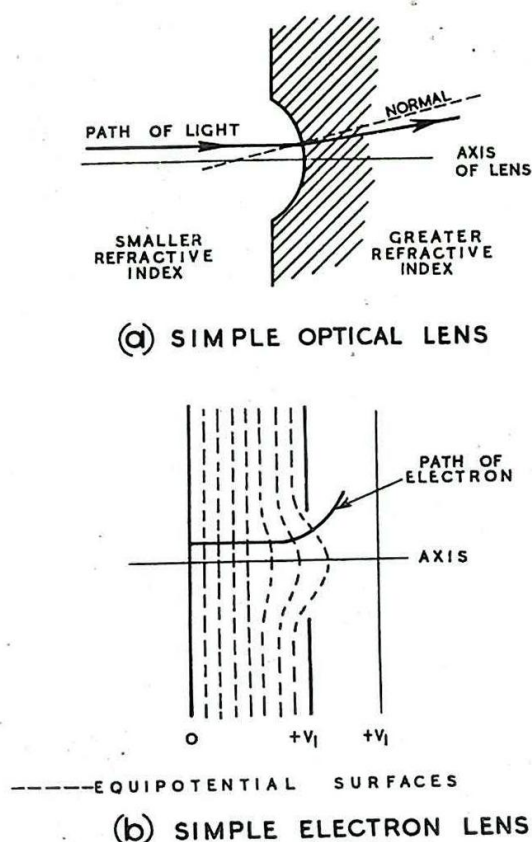


Fig. 267 - Simple lenses; (a) optical (b) electron.

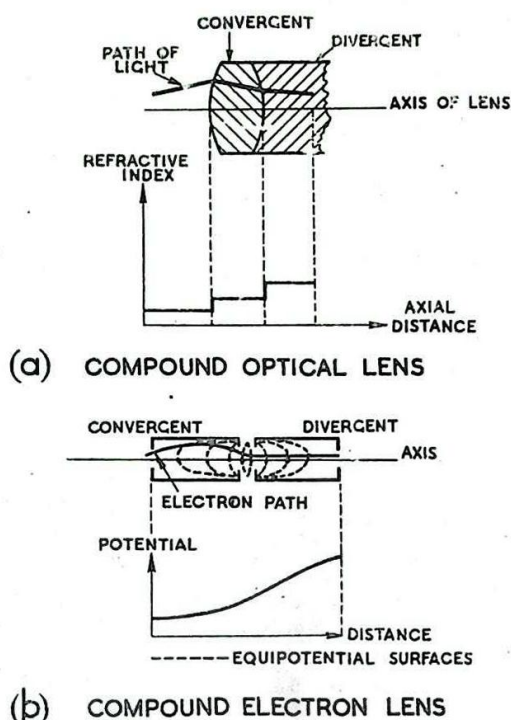


Fig. 268 - Compound lenses; (a) optical (b) electron.

A set of electrodes having rotational symmetry about an axis can be used to produce a symmetrical electrostatic field which has properties, in electron optics, somewhat similar to those of a lens in physical optics. Fig. 267 shows a simple type of electron lens which is produced by placing a plate with a circular hole in it between two other plates and then fixing the potentials of all three so that there are different potential gradients on the two sides of the centre plate. Fig. 268 shows a more complex type of electron lens. This lens is produced by two coaxial cylinders at different potentials. If the potential of the second cylinder is greater than that of the first, the equipotential surfaces are distributed as shown in the diagram. An electron which enters the first cylinder at a small angle to the axis experiences a force towards the axis while it is passing through the first cylinder and on to the second. As the electron enters and passes through the second cylinder it experiences a force away from the axis. The convergent angular deflection exceeds the divergent because the electron remains longer in the convergent portion of the lens, and therefore a given force can deflect it through a greater angle. The resultant converging action of an electron lens consisting of two cylinders depends on the dimensions and relative spacing of the cylinders and on the ratio of their potentials.

The simplest method of achieving electrostatic focusing of the beam in a CRT is shown in Fig. 269. Here the focusing system consists of two cylinders held at different potentials, both positive with respect to that of the cathode. The cylinder nearer the grid (First Anode) is commonly at a potential of the order of a quarter of

that of the other cylinder (Second Anode) with respect to the cathode. The point to which the electron beam is focused is determined by the ratio of the potentials of the first and second anodes. This ratio must be adjusted so that the focal point is brought to the fluorescent screen. This change in the ratio of the potentials on the two anodes is usually obtained by altering the potential of the first anode.

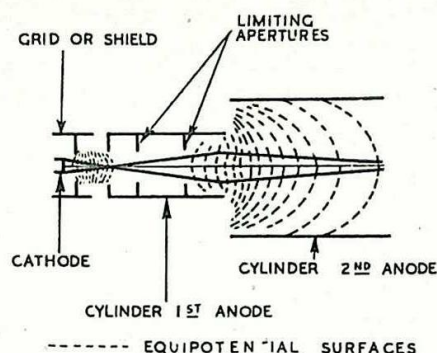


Fig. 269 - Two-anode focusing system of CRT.

With the two-anode system it is found that the controls of focus and brilliance are not independent. Change of voltage of the first anode (focusing) affects the intensity of the beam current (Sec. 6) and, therefore, the brilliance, whilst adjustment of the voltage of the grid alters the focusing to a small extent. This is to be expected when it is realised that any change of potential of any electrode will affect the potential ratios and accordingly modify the focusing of the electron beam.

When, as is usual in radar applications, it is required to produce a sharp spot whose focus is almost independent of the brightness-control, a three-anode focusing system is used (Fig. 270). It differs from the two-anode type by the addition of an extra anode on the side nearer to the grid. This new electrode is called the first anode and is commonly, but not always, kept at the same potential as the third anode, nearest the screen. The middle electrode, or second anode, is usually at an appreciably lower potential, about one-third or one quarter of that of the two other electrodes, relative to the cathode. Focusing is usually obtained by a variation of the potential of the second anode.

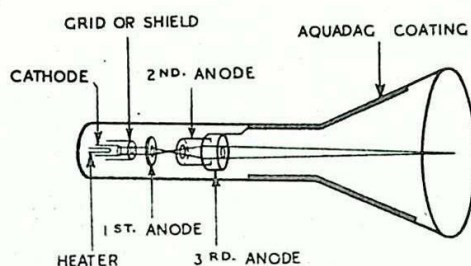


Fig. 270 - Three-anode focusing system.

The additional anode, between the grid and the focusing anode, can be compared to a certain extent with the additional grid which converts a triode to a tetrode. Variations in the electric fields on the side of the additional electrode remote from the cathode are thereby screened and have little effect on the field near the cathode. Thus, variation of the potential of the second anode of a focusing arrangement does not appreciably affect the intensity of the beam current, i.e. the brilliance. The presence of the first anode also minimises the small effect of changes of grid potential on the focusing system formed by the second and third anodes. A further advantage of the three-anode system lies in the relative independence of the deflection system on the focusing arrangements. In the two-anode tube variation of the first anode is liable to vary the velocity of the electrons between second anode and screen, and hence to affect the deflectional sensitivity.

The size of the spot obtainable by the above method is directly related to the accelerating potential used. If the potential difference between the third anode and cathode is low no amount of juggling with electrode design or relative potentials can produce a sharply focused spot. Typical potentials at the electrodes of a three-anode CRT, whose screen diameter is 12 inches, are as follows:-

Cathode	- 4000 V.
Control electrode	- 4010 to -4050 V. (variable brilliance).
Second anode about	- 3000 V. (variable focus).
First and third anodes	0 V.

13. Magnetostatic focusing

In magneto-statically focused tubes there are no focusing electrodes inside the tube. A steady magnetic field is directed along the axis of the tube by means of a coil, supplied with direct current (or possibly by means of permanent magnets) mounted round the neck of the tube beyond the anode, (Figs. 271 and 272).

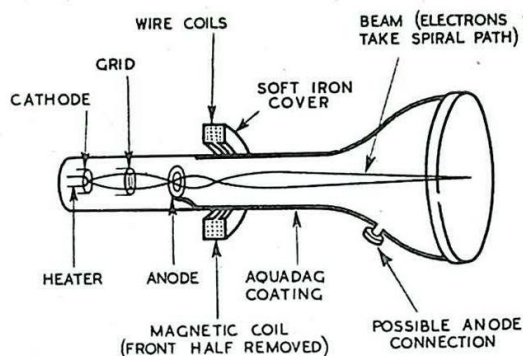


Fig. 271 - CRT with magnetostatic focusing.

Two methods are possible:-

- (i) Using a uniform axial field.
- (ii) Using a non-uniform field with large radial components.

Method (ii) is the one used in practice.

Consider an electron of charge e moving with velocity u at right angles to a magnetic field H . The electron will experience a force Heu at right angles both to the field and direction of motion (Fig. 272).

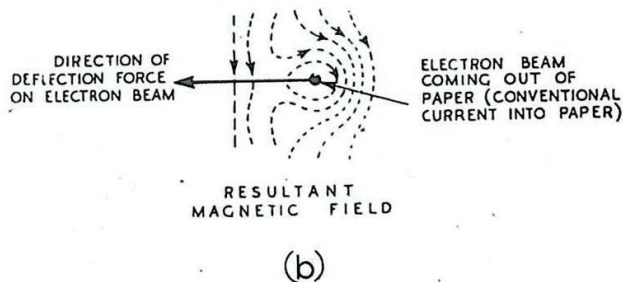
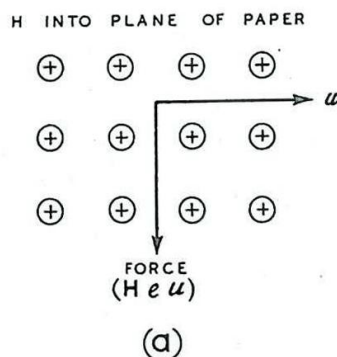


Fig. 272 - Force acting on an electron moving in a magnetic field.

This force may be considered as generated by the interaction of the applied field H and the magnetic field due to the electron current, as illustrated at (b).

It is this force on an electron travelling in a magnetic field which is utilised in magnetostatic focusing. The field used is axially symmetrical and is produced by placing a coil inside a metal cover with gap. When current flows through the coil the field produced is similar to that produced by the two annuli of north and south polarity; (Fig. 273). A and B are neutral points.

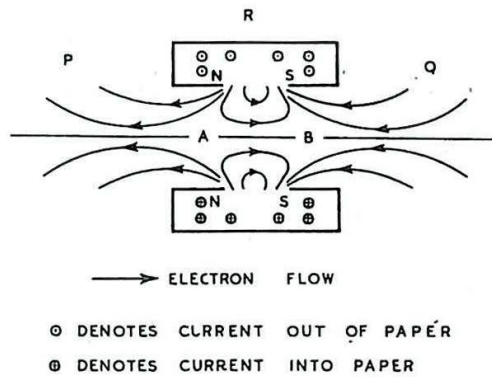


Fig. 273 - Magnetostatic field of focusing coil.

Before discussing the focusing action of the coil, we note that:-

- (i) in their travel from cathode to screen the electrons spend most of the time in space which is free from magnetic fields of significant magnitude.
- (ii) the electrons are moving with a high velocity along or very near to the axis. Hence only a radial component of the field will produce an appreciable deflecting force in such a direction as to cause the electron to spiral about the axis.

Observation of the lines of force reveals that there are two regions P and Q where the radial fields are relatively large. Near the axis the radial component is bound to be approximately proportional to the distance from the axis. The space R between P and Q contains a large axial magnetic field and near the axis this will be virtually constant. A simplification of such a magnetic field is shown in Fig. 274. A non-axial electron entering region P will start rotating about the axis.

When it leaves P it will have attained an angular velocity proportional to its distance from the axis. In the figure this velocity will be into the plane of the paper above the axis. In region R this velocity being perpendicular to the axial field will cause the electron to move towards the axis. Thus on leaving region R in addition to its rotation about the axis the electron will have a radial component of velocity. This radial component will be proportional to the angular velocity with which the electron entered the region R. Thus the radial velocity will be approximately proportional to the distance of the electron from the axis on entering P.

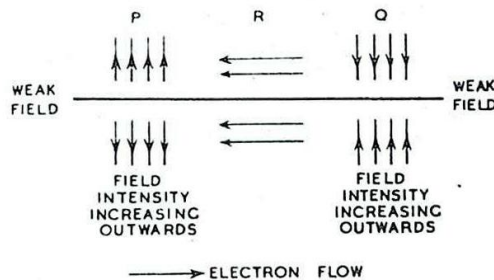


Fig. 274 - Idealised magnetostatic focusing field.

The electron now enters region Q. The radial field here can have no effect on the radial velocity but it will reduce the angular rotation; since the total inward flux at Q is equal to the total outward flux at P the ultimate angular velocity will be zero, but usually

rotation of the image exists. Thus an "off axis" electron leaves the field with a velocity towards the axis. This velocity is proportional to its distance from the axis on entering the field. It can be shown that, provided the disposition and intensity of the magnetic field are suitably adjusted, all the electrons diverging from a point on the axis outside the field will after passing through the field converge to another point on the axis.

Broadly speaking, at A of Fig. 275 there is twice the divergence at B, but the electron is twice as far from the axis so that the radial velocity after passing through the field will be twice as much.

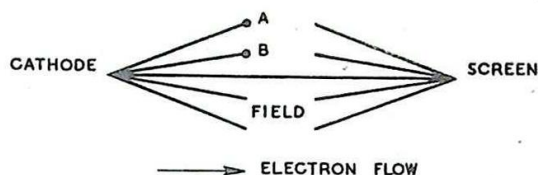


Fig. 275 - Direction of electron paths entering and leaving focusing field.

By suitable adjustment of the position of the coil and the current through it the focal point at which the electrons converge can be brought to the screen of the CRT.

14. Gas Focusing in Soft Tubes

A small quantity of inert gas such as argon or helium, admitted after the tube has been fully evacuated, provides the focusing action of the electron beam. The grid potential is usually about 50 volts below that of the cathode and practically all the emitted electrons go through the hole in the anode as a beam of small divergence. The anode potential is normally between 300 and 500 volts positive.

As the electrons pass along the tube from the grid, many collide with the gas molecules, and ionise them so that electrons are ejected and comparatively heavy and slow-moving ions are left behind. The positive ions form a core which exerts a considerable attractive force on the negative electrons. Provided the number of ions formed is sufficiently great the mutual repulsive force of the electrons is more than counterbalanced, and it is possible to converge the stream to give a spot on the screen of less than 0.5 mm. diameter.

The comparatively short life and dependence of focus on brilliance in soft tubes using gas focusing makes them unsuitable for most radar applications. Also, the inertia of the positive ions introduces a lag in the focusing action so that for rapidly varying deflection voltages the focus is impaired. Soft tubes, however, have the advantage that intensity of illuminations of the screen can be accomplished with a low accelerating potential. External power supplies are, therefore, simplified, and since the electrons are moving comparatively slowly, they can be deflected easily. Such advantages explain the use of soft tubes in some types of oscillograph, but even here they are being replaced by hard tubes using electrostatic or magnetostatic focusing.

DEFLECTION SYSTEMS

15. General

It is necessary that the electron beam should be deflected after it leaves the final anode. In its simplest form the deflection system is such that the beam of electrons can be deflected to and fro in a direction at right-angles to that of the beam.

There are two types of deflection systems, electric and magnetic; these systems are applicable to both soft and hard tubes.

16. Electric Deflection

An electron situated in a uniform electric field experiences a force in the direction of the field. Consequently a beam of electrons passing through a transverse electric field will be deflected. This deflection is usually obtained by applying a potential difference to a pair of parallel plates so situated within the tube that the beam passes between them shortly after it emerges from the focusing field; (Fig. 276). The beam experiences a force proportional to the field strength and therefore to the potential difference between the plates. It is bent towards the more positive of the two plates and the velocity component which the electrons acquire, at right angles to the axis of the tube, persists after they leave the deflector plates until they hit the screen. If the potential difference between the deflector plates is constant the beam is permanently deflected. Steady potential differences called Shift Voltages are used to adjust the initial position of the spot of light on the screen. If an alternating potential difference is applied to the plates, the spot is swept to and fro across the tube and, if the frequency is sufficiently great, persistence of vision will cause it to appear on the screen as a trace.

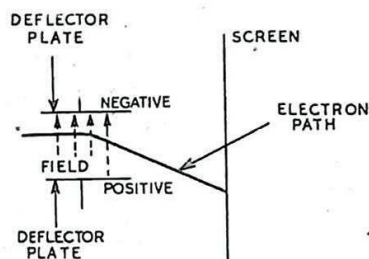


Fig. 276 - Deflection of electron by pair of parallel plates.

It is possible to obtain deflections in two perpendicular directions by mounting two mutually perpendicular sets of parallel plates in the neck of the glass envelope, (Fig. 277), one set of plates being mounted nearer the screen than the other. These two sets of plates are called the X and Y plates, to indicate the directions of the two deflections. To allow for the deflection of the beam, the plates are not usually parallel throughout their whole length but usually diverge towards the screen. The width of the second pair of plates (the pair nearer the

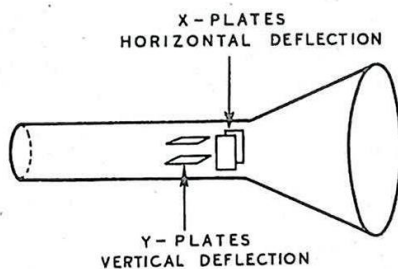


Fig. 277 - Deflector plate assembly.

screen. The width of the second pair of plates (the pair nearer the

screen) is larger than that of the first pair to allow for the deflection of the beam by the first pair. It is possible, by the application of suitably varying voltages to the two pairs of plates, to produce a trace at any desired inclination to the X and Y axes.

The deflection sensitivity of a CRT employing electric deflection is given by the deflection of the spot in millimetres for a potential difference of one volt between the two deflector plates. This deflection is found to be inversely proportional to the potential difference between the cathode and final anode, on the assumption that the mean potential of the deflector plates is the same as that of the final anode. Soft tubes are normally operated at much lower anode potentials than hard tubes, so their electric deflection sensitivity (about 0.5 mm/V.) is usually greater than that of hard tubes (about 0.2 mm/V.)

Derivation of Electric Deflection Sensitivity (Fig. 278)

Let:-

l be the length of the deflector plates,

t the distance between the plates,

s the length of the tube from the centre point of the deflector plates to the screen,

e the magnitude of the charge on an electron,

m the mass of an electron,

V_a the potential difference between the final anode and the cathode of the CRT,

V_p the potential difference between the plates,

u the velocity of an electron on entry.

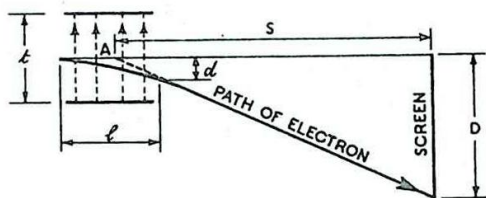


Fig. 278. - Electron undergoing electric deflection.

$$\text{The field-strength between the plates} = \frac{V_p}{t}$$

$$\text{The force on an electron} = \frac{V_p \cdot e}{t}$$

$$\text{Since force} = \text{mass} \times \text{acceleration} \text{ then the acceleration of the electron towards the positive plate} = \frac{V_p \cdot e}{t \cdot m}$$

$$\text{The time the electron is moving between the plates} = \frac{l}{u}$$

Since displacement = $\frac{1}{2} \cdot \text{acceleration} \times (\text{time})^2$ then the transverse displacement at the end of the plates is given by $d = \frac{1}{2} \cdot \frac{V_p \cdot e}{t \cdot m} \cdot \frac{l^2}{u^2}$

But the kinetic energy of an entering electron = $\frac{1}{2} mu^2$ = loss of potential energy as it passes from cathode to final anode = $e \cdot V_a$, so that $mu^2 = 2e V_a$.

$$\text{Therefore } d = \frac{l^2}{4t} \cdot \frac{V_p}{V_a}$$

Since the deflecting force is at right angles to the axis of the tube (initial direction of motion of electron) the electron path is parabolic and the point A, the apparent origin of the deflected beam, is halfway along the plates.

Therefore the displacement on the screen is given by

$$D = \frac{ds}{l/2}$$

$$= \frac{l^2}{4t} \cdot \frac{2s}{l} \cdot \frac{V_p}{V_a}$$

Then the displacement on the screen per unit of potential difference between deflector plates, i.e. the deflection sensitivity, is given by $\frac{D}{V_p} = \frac{l s}{2 \cdot t \cdot V_a}$

For a given tube and plate dimensions, the deflection sensitivity is inversely proportional to V_a (potential difference between the cathode and final anode). Since $V_a \cdot e = \frac{1}{2} mu^2$ the deflection sensitivity is inversely proportional to the square of the velocity (u) of the electron.

For a given value of V_a , the closer the plates, or the longer their length, or the longer the tube from the plates to the screen, the greater is the deflection sensitivity. The use of a long tube, however, introduces difficulties in producing a good focus. The focusing system has no control over the electron beam once it has left the system and, due to the mutual repulsion of the electrons, the more remote the screen the less sharp will be the focus. Hence a compromise is necessary between good focus and high deflection sensitivity.

17. Magnetic Deflection

It is often convenient to produce the deflection of the electron beam by a transverse magnetic field. This is produced by a pair of similar coils, one on each side of the neck of the glass envelope. These two coils are in series and are so wound that they produce a magnetic field in the same direction across the neck of the glass envelope. Alternative arrangements are shown in Fig. 279. The field is perpendicular to the direction of the electron beam so that the deflection is at right angles both to field and beam (See Sec. 13)

A steady current through the pair of coils produces a constant deflection of the spot; an alternating current sweeps the spot to

and free to form a trace. Since the deflection is a radial one, perpendicular to the direction of the field, the location of the spot on the screen depends on the position of the coils.

A second pair of coils at right-angles to the first pair deflects the spot in a direction perpendicular to that of the deflection due to the first pair of coils. (Fig. 280). By supplying each pair of coils with a suitably varying current a trace can be produced at any desired inclination to the X and Y axes.

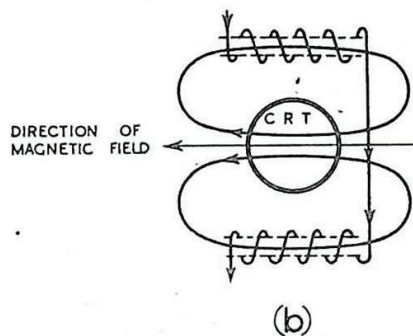
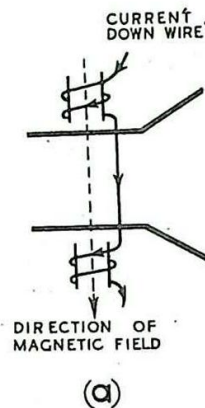


Fig. 279 - Alternative methods of producing magnetic deflection.

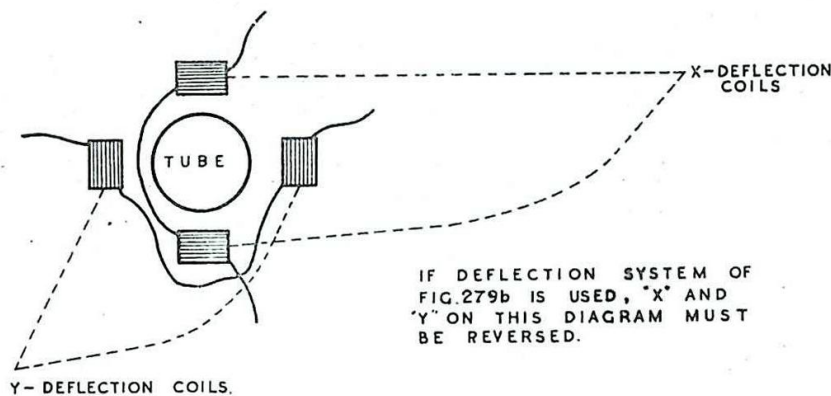


Fig. 280 - Coil assembly for producing magnetic deflection.

With magnetic deflection the deflection sensitivity is found to be inversely proportional to the square root of the potential difference between final anode and cathode. Thus the deflection sensitivity with magnetic deflection of soft tubes (low anode voltage) is greater than with hard tubes (high anode voltage) but the difference is not so marked as with electric deflection where the deflection sensitivity is inversely proportional to the potential difference

between cathode and final anode.

Derivation of Magnetic Deflection Sensitivity (Fig. 281)

If an electron moves into a uniform magnetic field so that its direction of motion is at right angles to the field, it will experience a force perpendicular both to the field and to the motion.

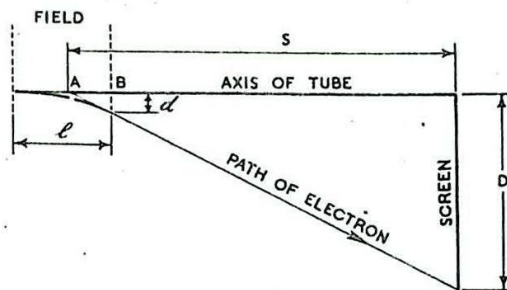


Fig. 281 - Electron undergoing magnetic deflection.

Let m be the electronic mass,

e be the electronic charge,

H be the magnetic field strength,

V_a be the potential difference between final anode and cathode of the CRT,

l be the extent of the magnetic field (assumed uniform),

s be the distance to the screen from A, the apparent origin on the axis of the deflected beam; (this is approximately the centre of the field),

u be the electron velocity.

Then the force on the electron is Heu ,

so that the acceleration is $\frac{Heu}{m}$.

If this force acts for a time t , this causes the electron to be deflected a distance $d \doteq \frac{1}{2} \left(\frac{Heu}{m} \right) t^2$.

$$\begin{aligned} \text{Since } t &= \frac{l}{u}, \text{ this gives } d \doteq \frac{1}{2} \frac{Heu}{m} \frac{l^2}{u^2}, \\ &\doteq \frac{Hel^2}{2mu}. \end{aligned}$$

The deflection D of the spot on the screen is then given by $D \doteq \frac{s}{AB} d$.

$$\begin{aligned} \text{If we take } AB &\text{ approximately equal to } \frac{l}{2}, \text{ this becomes } D \doteq \frac{s}{l/2} \frac{Hel^2}{2mu} \\ &\doteq \frac{lsHe}{mu} \end{aligned}$$

Writing $\frac{1}{2} mu^2 = eV_a$, we have

$$u = \sqrt{\frac{2eV_a}{m}},$$

so that $D \doteq l_s H \sqrt{\frac{e}{2m V_a}}$.

Hence the deflection sensitivity, $\frac{D}{H}$ is given by

$$\frac{D}{H} = l_s \sqrt{\frac{e}{2m V_a}}.$$

For given tube dimensions and arrangement of coils the deflection sensitivity is inversely proportional to the square root of V_a . Alternatively it is inversely proportional to the velocity of the electron.

For a given V_a , the longer the tube from the coil to the screen or the more extensive the field, the greater is the deflection sensitivity. The use of a long tube, however, leads to a deterioration of focus.

In practice magnetic deflection sensitivity is often measured in millimetres per milliamp of current through the coils instead of per unit of magnetic flux.

DISTORTIONS AND THEIR CORRECTION

18. Trapezium Distortion

Trapezium distortion can occur in all CRTs in which electric deflection is utilised, but it is more apparent with hard tubes since in these much higher voltages must be applied to the deflector plates to produce full scale deflection. This distortion arises because the potentials of the deflector plates nearer the screen, say, the X-plates may affect the deflection sensitivity of the other pair, the Y-plates.

If the movements of the spot fill the face of the tube as in television and in certain radar displays the effects of trapezium distortion are most noticeable. The rectangular picture (or Raster) is altered to a trapezium, shown dotted in Fig. 282, thereby giving the name to this type of distortion.

It has been thought that trapezium distortion is due to the interaction between the electric fields of the two pairs of plates. However, it can be shown that it can be produced if only one pair of plates is used, the deflection at right-angles being caused by magnetic means. If the electric deflection system is nearer the screen than the magnetic deflection system trapezium distortion can occur.

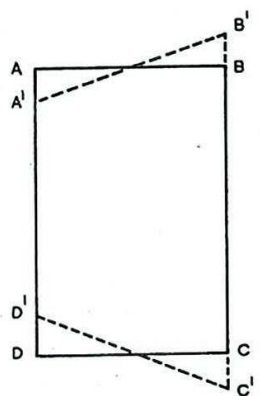


Fig. 282 - Trapezium distortion.

Consider the Y-deflection to be produced magnetically. Suppose a positive potential of, say, 500 V. is applied to the upper

X-plate (Fig. 283(a)), the lower plate being earthed, as are the final anode and glass envelope. The equipotential surfaces will be distributed in a manner similar to that illustrated. Now consider them cut by a plane mid-way between the X-plates and containing the beam, which has already been deflected magnetically in the Y-plane. This cross section is illustrated at (b). The field distribution acts as an electronic lens (see Sec. 12), the beam being deflected towards the normal to the equipotential surfaces on entering (accelerating field) and away from the normal on leaving (retarding field). Thus in the case illustrated the Y-deflection is decreased by the X-plate potential distribution. Similarly it may be shown that if one of the X-plates is negative with respect to the final anode the Y-deflection is increased.

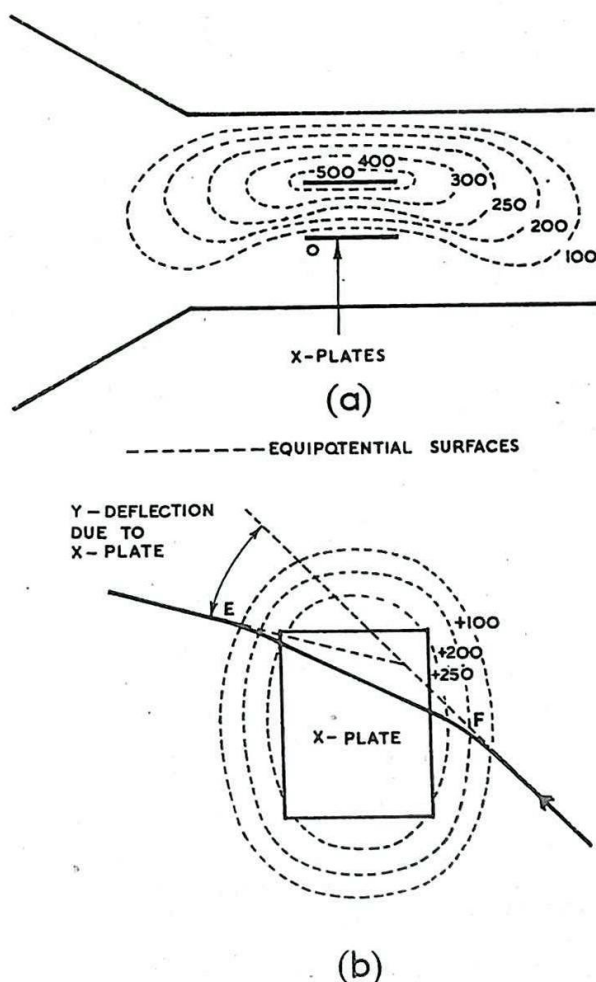


Fig. 283 - Cause of trapezium distortion.

If electric deflection is used for both X- and Y-displacements, the application of deflecting potentials to the Y-plates modifies the appearance of the field produced by the X-plates; but even when this action is taken into account, the lens field produced by the X-plates still alters the Y-deflection in a manner similar to that described above.

One method of reducing trapezium distortion is to use a balanced deflection voltage for the pair of plates, say X-plates, nearer the screen. If the deflecting potential applied to one plate at any instant is equal in magnitude and opposite in sign to that applied to the other, the potential of the space midway between the plates in the path of the beam is approximately constant and the Y-deflection is not altered.

Alternatively deflector plates corrected for trapezium distortion may be used (e.g. if the tube is used in a double-beam

oscilloscope). The pair of deflector plates nearer the final anode is so assembled that the plate separation is not constant: that is, the plates are not parallel. Hence the deflection sensitivity depends upon whether the beam traverses the region between these plates at a point where the plate separation is small or great. In the former case the sensitivity is greater than in the latter. So if the plate separation is correctly tapered, compensation for trapezium distortion can be achieved.

Another method is to shape the plates nearer the screen (X-plates) so that these plates produce a deflection which has a component in the Y direction so that the trapezium distortion effect is counteracted. This method is illustrated in Fig. 284(a), where the X-plates are shown to be curved. The arrows indicate the direction of deflection produced by the X-plates. A practical example of shaped X-plates is shown in Fig. 284(b); the equipotential lines (shown dotted in the diagram) are curved and the effects are similar to those obtained with curved X-plates.

It should be noted that if this method for correcting trapezium distortion is employed it operates only if the deflecting voltages applied to the plates nearer the screen are unbalanced and also only if the correct plate is earthed (connected to the final anode).

A more satisfactory method for correction of trapezium distortion is shown in Fig. 285. By shaping the X-plates as shown the equipotential lines at F (Fig. 283(b)) can be made concave as seen from the final anode. Then if the potential distribution is such as to cause the beam to be deflected towards the normal at F

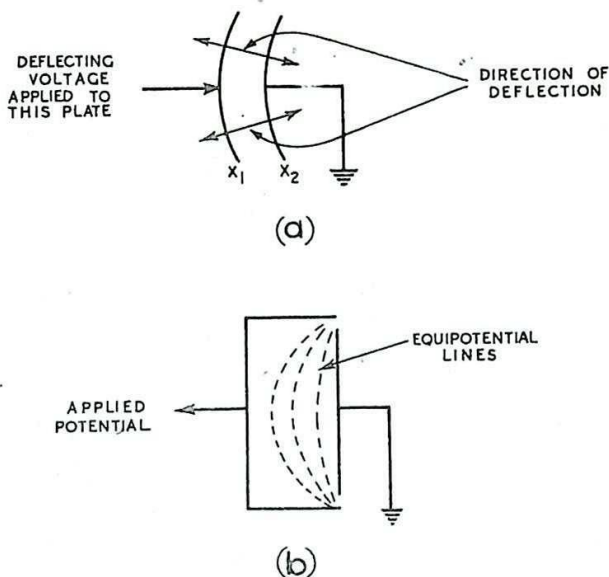


Fig. 284. - Shaped plates to eliminate trapezium distortion.

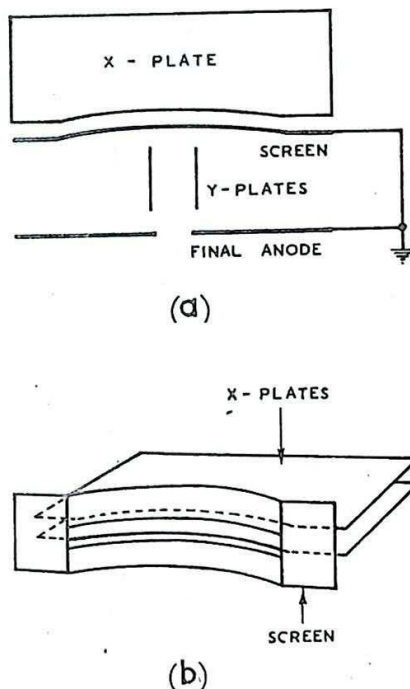


Fig. 285 - Alternative method for correcting for trapezium distortion.

it will be deflected away from the normal at E, so that the deflection introduced is outwards at F and inwards at E. A similar argument holds if the field at F is retarding instead of accelerating. The opposing deviations of the beam at F and E can now be balanced by correct shaping of the plates and trapezium distortion can be avoided. The shaped and slotted screen, which is connected to the final anode and placed between the X- and Y-plates (Fig. 286) is necessary; otherwise the equipotential lines tend to bulge into the space between the Y-plates. This correction method does not affect the symmetry of the X-plates and balanced deflection potentials may be applied or alternatively either X-plate can be earthed.

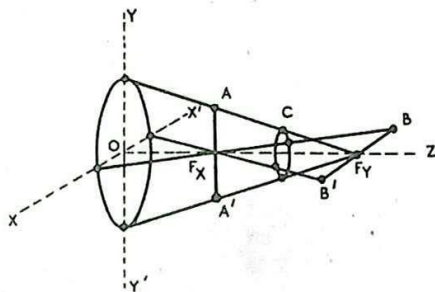


Fig. 286 - Astigmatism produced by optical lens.

With magnetic deflection, if the field due to a pair of deflector coils is not uniform across the neck of the CRT, the distance the spot moves when on one side of the screen will not be the same as when it is on the other. Such an arrangement will produce a distortion similar in appearance to trapezium distortion.

19. Deflection Defocusing

The speed of the electrons in the region of the deflection system depends upon the potential difference between the deflection system and the cathode. It is important that this should remain constant since a variation in speed would move the focus off the screen. If unbalanced voltages are applied to the deflector plates the mean potential of the deflection system does not remain constant and defocusing may occur. Thus if the focus is correctly adjusted at the centre of the screen it deteriorates as the distance from the centre increases. This phenomenon is called deflection defocusing and may be avoided by the use of balanced deflection voltages or by magnetic deflection.

20. Astigmatic Distortion

In order to understand the meaning of the word astigmatism consider first the focusing of an ordinary light beam by passage through a convergent lens. (Fig. 286). If the lens is symmetrical about the axis OZ the amount of convergence is equally great in all axial planes, and a sharp spot of light is produced in the focal plane. If the convergence is a maximum in the plane XX'Z and a minimum in the plane YY'Z then the beam will be converged to a line focus at two positions AA' and BB', these lines being in mutually perpendicular directions. The best focus obtainable is at C where the beam is converged to a circular focus, but of finite size. This type of defect often occurs in the lens of the eye and gives rise to the defect of vision known as astigmatism.

In a cathode ray tube it sometimes happens that the electric lens is asymmetrical due to misalignment of the electrode system

in the tube. When this occurs it is found that manipulation of the focusing control produces a sharp line on the screen in one or other of two perpendicular directions, or a circle of finite size, but never a small sharply defined spot. This fault is said to be due to the presence of astigmatism in the tube.

Astigmatism can be corrected by having the mean potential of each pair of deflector plates separately adjustable at a value other than that of the final anode. This mean potential of a pair of plates should be independent of shift voltages and variable deflection voltages. Such an arrangement introduces deliberate distortion into the electric fields between the electrodes so as to counterbalance the distortion producing the astigmatism.

21. Bulb Charge and Deflector Plate Current

When the main electrons in the beam strike the screen they eject secondary electrons from it and it is these secondary electrons travelling to the anode which make the major contribution to the anode current. The potential of the screen becomes negative, to that of the anode by an amount sufficient to ensure that equilibrium conditions are set up, i.e. the rate at which secondary electrons leave the screen to go to the anode is the same as the rate of arrival of the primary electrons at the screen. The potential difference developed is of the order of 100 volts.

With some tubes the screen picture disappears if the glass of the screen is touched. This minor defect is due to the induction of a large negative charge on the screen, sufficient to repel the beam electrons so that they never reach it. The picture will reappear after a short time, working its way back from the edges of the screen inwards, as the charge leaks away.

The electrons returning from the screen should pass to the final anode which is normally at earth potential. If, however, the steady potential of a deflector plate is positive with respect to earth, electrons are drawn to this plate. This effect may be likened to the flow of grid current in the input circuit of a valve amplifier (Chap. 9, Sec. 3). The non-linearity which is introduced in either case by the flow of current does not produce appreciable distortion unless the input resistance of the deflection system or valve circuit is sufficiently small as to be comparable with the output resistance of the source of voltage.

Errors due to deflector plate current distortion can therefore be reduced by feeding the deflector plates from a source of low output resistance, e.g. from a cathode follower. The electron current drawn by the deflector plates can be minimised by constructing the electrodes in such a manner that the deflector plates are shielded by the final anode, or by arranging that the maximum potential applied to any deflector plate does not rise appreciably above final anode potential (this latter method is not usually practicable).

22. Stray Fields Leading to Distortion

Alternating magnetic fields from components such as transformers and chokes act on the electron beam and may lead to an oscillatory motion of the spot so as to draw it out into a line or ellipse. The effect can be distinguished from normal defocusing because it is not removed by alteration of the focusing controls. In some cases the earth's magnetic field may introduce a deflection which varies in magnitude and direction as the position of the tube is changed relative to the earth's field. The usual way of overcoming the effect of stray magnetic fields is to surround the tube

with a mu-metal shield. If the shield is also earthed it protects the beam from any stray electric fields which may be set up by nearby apparatus.

23. COMPARISON OF FOCUSING AND DEFLECTION SYSTEMS

For small tubes with screens up to about 5 in. in diameter, it is usual to employ electrostatic focusing and electric deflection systems. This arrangement is convenient since it avoids the use of deflection equipment outside the tube. If electric deflection is used with larger tubes, involving high accelerating voltages, it is necessary to place the deflector plates of each pair close together in order to obtain reasonable deflection sensitivity. However, if the plates are too close together there is a danger that the beam of electrons will strike the edge of a plate before full deflection can occur so that the beam will be cut off from the screen. This effect can be minimised by increasing the plate separation towards the screen, but this arrangement has limitations and the method of deflection by magnetic fields is probably preferable.

When magnetic deflection is used it is important that the deflecting field, over the area of cross-section of the neck of the tube, be uniform in order to avoid errors of defocusing. One way of ensuring that the field is as uniform as possible is to make the area of cross-section as small as possible. If this is done it is often convenient to use magnetostatic rather than electrostatic focusing since the latter involves the insertion of a complex electrode system inside the neck of the tube. There is always the danger that the electrodes in an electrostatic focusing system may be misaligned. This fault will give rise to astigmatism which can usually be corrected, but only at the expense of fairly complex external circuits. A magnetostatic focusing system can be adjusted in position to avoid such astigmatism. However, it is not a simple matter to adjust a focus coil in position so as to be sure of obtaining the best possible focus attainable, and in spite of the possible disadvantages outlined above electrostatic focusing systems are often preferred.

24. POST-DEFLECTOR ACCELERATION

To avoid the difficulty, inherent in the normal hard tube, of poor deflection sensitivity, tubes are sometimes used in which a large part of the acceleration of the electrons takes place after they have been deflected. It is then possible to arrange that the beam is deflected when at low velocity so as to give a good deflection sensitivity, and yet for the beam to have a large velocity when it strikes the screen so that a bright trace is obtained. A tube of this type is shown in Fig. 287 and involves the use of an extra anode in the form of a narrow ring of aquadag coating on the glass envelope

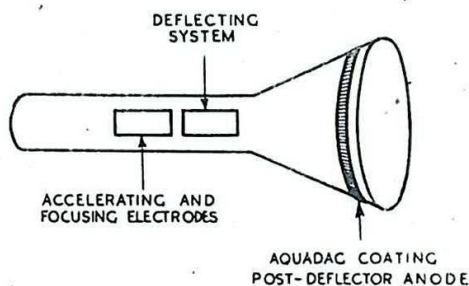


Fig. 287 - Extra electrode assembly for post-deflector acceleration.

just before the screen. This anode is usually maintained at about twice the potential of the last anode of the normal accelerating system. A deflection sensitivity 3 to 5 times as great as with an ordinary type can be obtained in this way. This method of working is called Post-Deflector Acceleration.

25. DOUBLE-BEAM TUBE

This is a tube of special construction in which the beam, after leaving the final accelerating anode, impinges on the edge of a splitting plate placed midway between the Y-plates. The splitting plate is connected to the final anode which is at earth potential. The beam is thus split into two sections (Fig. 288) each of which can be deflected almost independently in the Y-direction by voltages applied between Y_1 and earth and Y_2 and earth. As might be expected, the potentials applied to, say, the Y_1 plate affect to some extent the Y_2 trace and vice versa. A positive voltage at the Y_1 -plate produces an upward deflection of the spot whilst at the Y_2 -plate it produces a downward deflection. Both beams traverse the X-plates in the usual way, and are deflected simultaneously in the horizontal direction. This tube is of particular use in an oscillograph where it is desired to examine two potentials simultaneously.

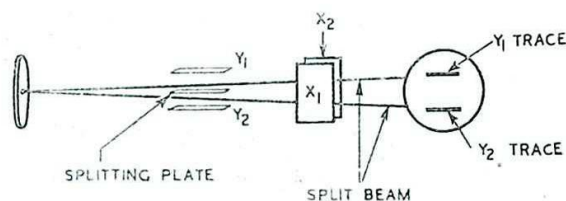


Fig. 288 - Electrode assembly to produce double (or split) beam.

It is possible by the use of external circuits to produce a double trace on an ordinary tube. Here the X-movement of the spot is made to occur successively first across, say, the lower half of the screen, and then across the upper. In this case corresponding vertical displacements on the two traces do not occur simultaneously.

POWER SUPPLIES AND SHIFT NETWORKS

26. General

Steady potentials are required for the various electrodes of a cathode ray tube. Thus, for a three-anode tube with electrostatic focusing, steady potentials are needed at the cathode, grid and the three anodes. A tube with magnetostatic focusing normally requires steady potentials for cathode, grid and one anode, and also a supply of direct current for the focus coil. Electric deflection

involves steady potentials for use as shift voltages so that the picture traced on the CRT screen may start from a convenient point. Magnetic deflection needs direct currents for producing shifts.

27. Power Supplies

The power supply for an electrostatically focused tube should be capable of supplying high voltages of the order of 5000 volts for a large tube, but need supply only a small current. The total load current taken by the CRT is usually less than $\frac{1}{2}$ mA. Half-wave rectification is therefore all that is necessary, and the smoothing circuit of the power unit can employ a resistor, rather than a choke. The high voltages required for the larger CRTs may be obtained by means of voltage-doubler circuits.

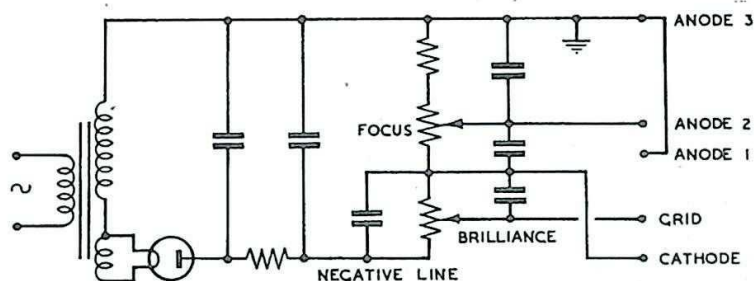


Fig. 289 - Supply circuit for three-anode CRT.

It is advisable to have condensers across all points of supply which are connected to the tube, otherwise a voltage ripple on the supply might be communicated to the electrodes so as to cause variations in brilliance or focus of the trace. A typical supply circuit, suitable for a three-anode electrostatically focused CRT (or for the anode, cathode and grid of a tube with magnetostatic focusing) is shown in Fig. 289.

As shown in Fig. 289 the third anode is at earth potential whilst all other electrodes shown are at negative potentials. This is because the deflector plates in a tube with electric deflection are connected to circuits in which it is convenient (but not essential) to have them at or about earth potential. In order to avoid strong electric fields between these plates and the final anode, this anode itself must be at earth potential.

Another reason for earthing the final anode is that the potential of the fluorescent screen is usually about 30 to 100 volts negative with respect to that of the final anode. If the cathode is earthed the screen is thereby maintained several hundreds or thousands of volts above earth so that strong electrostatic fields may exist between the screen and neighbouring conductors. This can

result in extraneous deflections of the spot if these conductors are moved. If the anode is earthed the cathode and its associated heating circuit must be well insulated from earth.

28. Shift Voltage for Electric Deflection Systems

The shift voltages are usually, but not always, applied to a pair of plates so that the mean potential of these two plates is the same as that of the final anode of the tube. This arrangement avoids trapezium and defocusing distortions, provided that any fluctuating voltages applied to the plates are also balanced. The shift voltages may be obtained either from the power supply which provides steady potentials to other electrodes of the tube (Fig. 290) or from a separate power supply. One deflector plate of a pair is always maintained as much positive, compared to the final anode potential, as the other is negative, so that the mean potential of the pair of plates is that of the final anode. Condensers C bypass to earth any ripple on the shift voltage supplies.

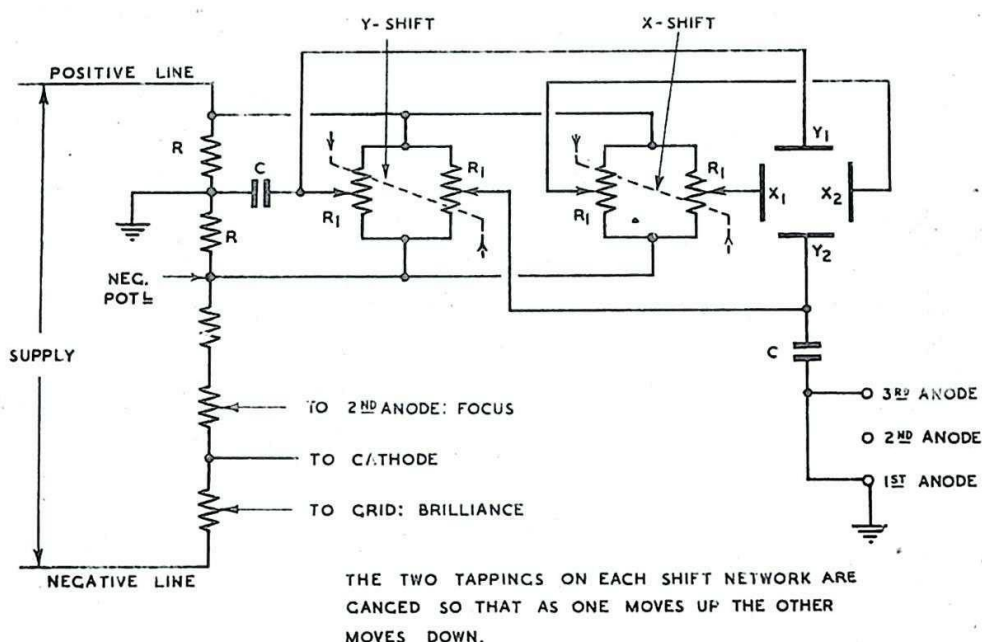


Fig. 290 - Supply for CRT showing production of shift voltages.

Each deflector plate must be connected, usually through a resistance of about two megohms (R and part of R_1 in Fig. 290) to the final anode (earth potential); this connection fixes the mean level of the deflector plates and is analogous to the use of a grid leak for determining the bias level in a valve amplifier circuit.

It is common in radar applications, involving the use of large CRTs to adjust the mean potential of each pair of plates with respect to the anode so that astigmatic distortion may be eliminated. The mechanism which provides this adjustment is called the Stig Control. A method is shown in Fig. 291.

A separate similar arrangement provides shift voltages and astigmatic correction for the X-plates.

In the figure, R_2 and R_3 are similar variable resistors forming the Stig Control so ganged that as R_2 is decreased in value R_3 is increased by the same amount, and vice versa. Thus, the mean value of the potential on a pair of plates is raised or lowered above or below earth potential (final anode potential).

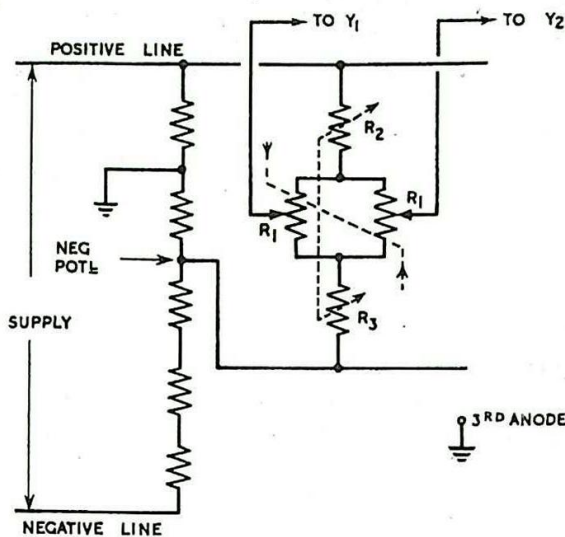


Fig. 291 - Supply network showing Y shift and Y stig. control.

A simple type of shift network which is sometimes used with small CRTs is shown in Fig. 292. Here no attempt is made to keep the mean potential of the deflector plates the same as that of the anode, nor is any attempt made to correct for astigmatism. It is also assumed that the power supply is being used for other sections of radar equipment, so that it is convenient to earth its negative line.

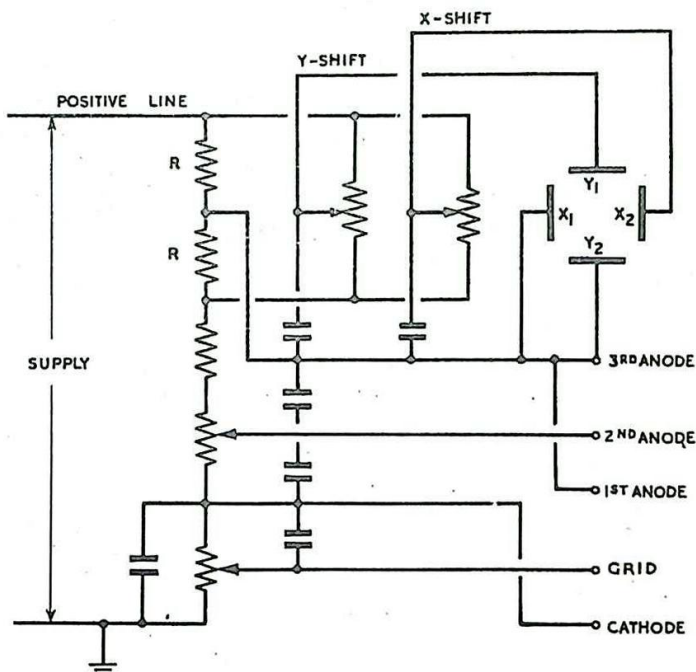


Fig. 292 - Simple type of shift network.

29. Shift Current for Magnetic Deflection Systems

A direct current through the deflector coils is necessary to obtain shift when magnetic deflection is used on a CRT. In most applications it is not essential that the shift should be made to work both ways and a simple arrangement such as is shown in Fig. 293 is adequate.

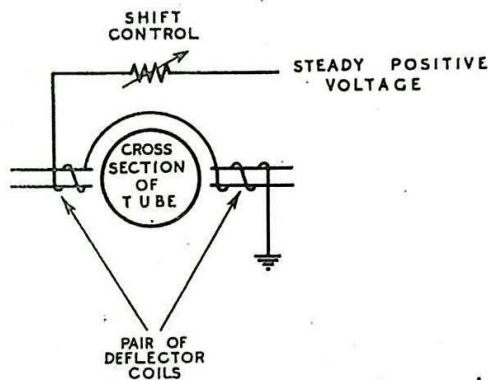


Fig. 293 - Shift control for magnetic deflection.

INPUT CIRCUITS

30. Input Circuits to Deflector Plates

The arrangements for applying input voltages to the deflector plates depends on the source of the input, and on whether shift voltages are also required. An invariable rule is that the mean potential of each pair of plates must be near that of the final anode. Fig. 294 shows a simple arrangement with no shift voltages; C isolates plate Y_1 from any steady voltage (which would produce shift) in the input supply, whilst resistor R (1 to 2 megohms) provides a path back to the final anode for electrons collected by Y_1 from the beam.

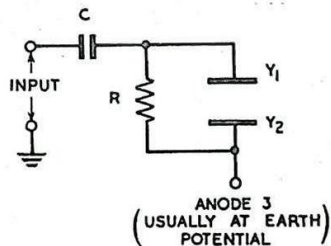


Fig. 294 - Input circuit to deflector plates.

Since in the absence of deflector plate current the resistance between a pair of deflector plates is of the order of 20 megohms, the effective load across which the input is developed is R (1 to 2 megohms).

The fact that the CRT has such a high input resistance is of importance since it means that its connection across a circuit does not usually disturb the conditions of that circuit appreciably. The inter-plate capacitance, of the order of 20 to 50 pF, is, however, shunted across the circuit supplying the plates and may cause a distinct modification of the input voltage.

If it is desired to apply both shift and work (input) voltages to one plate the circuit of Fig. 295 may be used. The resistor R is taken to the shift voltage tapping instead of directly to the anode. C_1 bypasses to anode, and so to earth, any ripple that there might be on the shift supply; the tapping on the shift slider is, therefore, at a constant potential. Thus, R is the effective load on the input source across which the work voltages are developed.

Fig. 296 shows the arrangement whereby a work and shift voltage can be applied to one deflector plate and an equal and opposite work and shift voltage can be applied to the other plate. Such an arrangement maintains the mean potential of the pair of plates at a value always equal to that of

the anode, and so avoids trapezium and deflection defocusing of the trace. A similar circuit allows two distinct inputs to be applied simultaneously, when the resultant deflection will be at any instant proportional to the difference of the two input voltages.

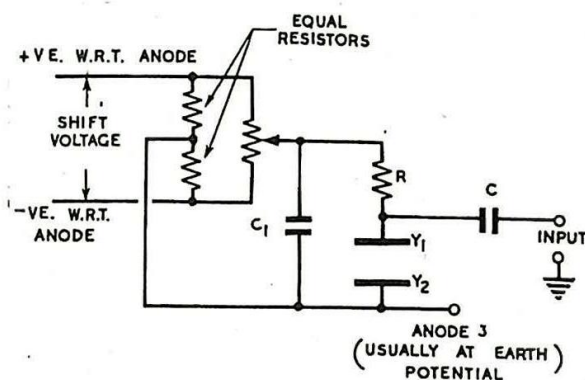


Fig. 295 - Input circuit when shift and work voltages are applied to same plate.

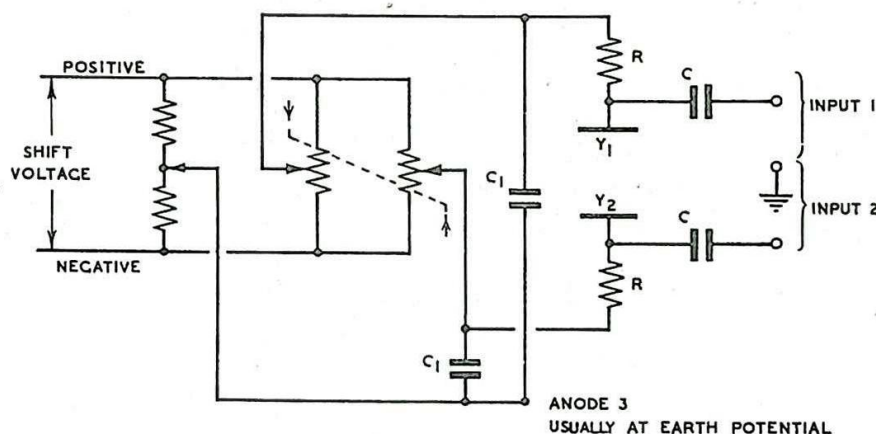


Fig. 296 - Input circuit for balanced input and deflection voltages.

31. Input Circuit to Grid and Cathode

It frequently happens that it is desired to black-out the spot of light on the screen of a CRT during some portion of its work cycle. For example, when a trace is being moved to and fro across the screen it may be necessary to eliminate the fly-back so as to prevent its obscuring the forward trace. Alternatively, it may be necessary to brighten the spot over some portion of the trace (intensity modulation). These effects can be secured by applying a pulse of voltage to the grid (or cathode). A rise of potential at the grid or a fall of potential at the cathode increases the brilliance.

Fig. 297 shows how the additional positive or negative pulses may be applied to the grid and cathode of the CRT. R_1 is the load across which the grid input voltage is developed, whilst

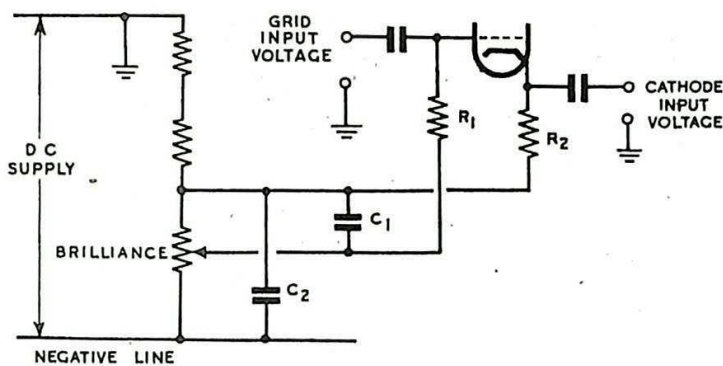


Fig. 297 - Combined grid and cathode input circuits.

R_2 is the load across which a cathode input voltage would be applied. C_1 and C_2 bypass the ripple on the supply voltage and ensure that one end of each of the resistors R_1 and R_2 is at a steady potential. Fig. 298 shows the circuit arrangement if pulses are to be applied to the grid alone, while Fig. 299 shows the arrangement if pulses are to be applied to the cathode alone. These circuits are in common use in radar equipments. In each case the condenser C must be capable of withstanding a high voltage, since one plate is normally connected to a large negative potential (potential of grid or cathode) whilst the other plate (to which the pulses are applied) is usually at a small positive potential.

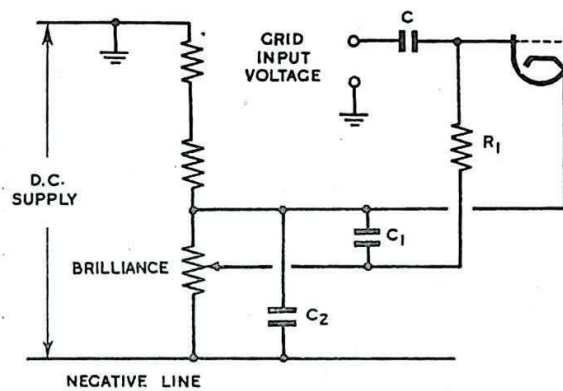


Fig. 298 - Grid input circuit.

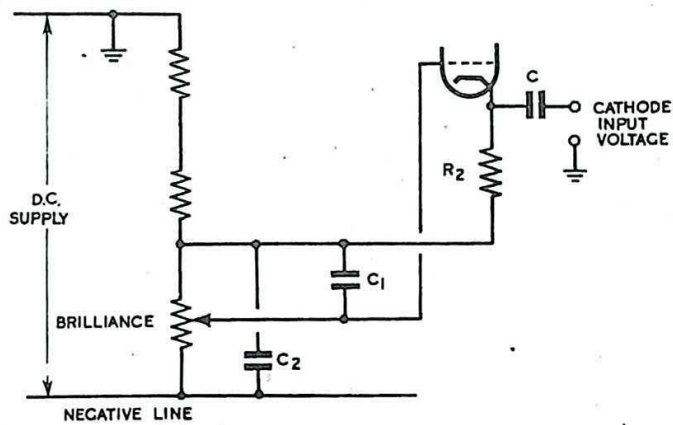


Fig. 299 - Cathode input circuit.

HARD VALVES

32. Introduction

In the following sections are discussed some of the special problems which apply to the use of hard valves in circuits peculiar to radar, as distinct from conventional radio circuits. No attempt will be made to describe the more normal uses of valves, which are adequately dealt with in other works.

33. Power Valves for Radar Transmitters

In radar systems which transmit pulses of short duration, of the order of a microsecond, at intervals of about one millisecond, the problem of power conversion is somewhat different from that which arises in conventional radio transmission. The transmitting valve, whether a triode such as might be used at 50 Mc/s., or a centimetric oscillator, such as a magnetron, is required to emit high-voltage pulses which, if they were allowed to last for more than a few microseconds, would rapidly destroy the valve due to excessive anode dissipation and consequent overheating. Owing, however, to the relatively long quiescent interval between pulses, the valve is able to recover from its temporarily overworked condition. In this way power valves may be used without particularly elaborate cooling systems, at voltages far above their ordinary CW rating. Such a valve must have a very high emissivity, and therefore a large cathode, compared with that of a radio transmitting valve, to generate adequate peak currents. The anode, however, need not be unduly large, and air-cooling is usually employed. In some cases the anode takes the form of a brass block which is cooled by air pumped through cooling ducts. In others, fins in a current of cool air provide sufficiently rapid heat dissipation.

The precautions usual in high-frequency circuits must be taken in the construction of the valves. Electrode leads must be kept as short as possible and inter-electrode capacitances reduced to a minimum. Acorn and Door-Knob valves have been designed to meet these requirements, and may be employed as radar transmitting or receiving valves at frequencies up to about 600 Mc/s. Above this frequency special types of transmitting valves must be used in which the tuned circuits form an integral part of the valve itself; examples of these are given in Chap. 8 Sec. 17. The limitations of conventional valves at high frequencies are further discussed in Chap. 7 Secs. 24. and 25.

34. Suppressor-Grid Characteristics

In many radar control and generating circuits multi-electrode valves are employed under conditions not normally encountered in conventional radio practice. In particular the use of the suppressor grid of a pentode as a control electrode is common. A brief description is given below of some of the effects of varying the suppressor-grid potential.

Since the suppressor grid was originally introduced between screen grid and anode to eliminate the well-known kink in the $I_a - V_a$ characteristics of a tetrode it is to be expected that if the potential of the suppressor grid is raised substantially above that of the cathode the kink in these curves will reappear. This does, in fact, happen, and if the suppressor grid is not held at a low potential the valve acts like a tetrode. Over the region of "negative slope resistance" an increase in anode voltage causes a decrease in anode current and an increase in screen current.

If the suppressor grid is made negative with respect to cathode the anode and screen-grid currents are usually very much affected. A reduction in the suppressor-grid potential, making it more negative, increases the potential barrier between screen grid and anode, reduces the anode current and increases that of the screen. Normally the suppressor grid must be made very negative before the anode current is completely cut off. For most RF pentodes with potentials at the anode and screen grid of a few hundred volts the suppressor grid must be made from 50 - 100 volts negative to cathode before anode current is cut off. Some valves have been specially constructed with closely wound suppressor grid wires, so that cut-off of the anode current occurs for only a few volts of negative bias on the suppressor. These valves may be especially suitable for use as Gate Valves or in Transitron oscillators or relays.

In a Gate Valve signal voltages are usually applied to the control grid, whilst gating pulses are applied to the suppressor grid or some other electrode. Generally the suppressor grid is held at a negative potential, sufficient to cut off anode current. Positive-going pulses are applied, raising the suppressor-grid voltage sufficiently to allow anode current to flow (provided the control grid is above cut-off potential). Thus only those signals are effective which appear at the control grid during the intervals of the positive-going pulses at the suppressor.

In the Transitron circuits use is made of the effect of the suppressor grid on the screen current.

Fig. 300 shows dynamic characteristics of two valves CV1065 and CV1091, operating under the conditions shown in Fig. 301. In both cases the screen-grid voltage is seen to depend on that of the suppressor grid in a curious manner. Consider the curve marked ABCD in Fig. 300. From A to B the suppressor-grid voltage is below the level sufficient to cut off anode current, and there is little change in screen current as the suppressor voltage is varied over this region. At B anode current begins to flow, and for the region B to C an increase in suppressor potential increases I_a and decreases the screen current I_s , so that the screen voltage rises.

From C to D the screen current increases, as the suppressor voltage rises, probably due to secondary emission from the anode.

As indicated in the diagrams, this portion of the screen-voltage, suppressor-voltage dynamic characteristic is very pronounced for low values of anode potential only and corresponds to the kink which is introduced in the $I_a - V_a$ characteristic of the pentode, due to the positive suppressor-grid voltage. Some secondary emission also occurs from the suppressor, and if the voltage at this electrode is raised for a short interval sufficiently far above cathode potential secondary emission may follow at such a rate as to cause the suppressor grid to remain at a high potential, if the path connecting it to the cathode is not of very low resistance.

If screen and suppressor grids are connected by a condenser, as in transitron circuits, the region BC of the above curve corresponds to a region of regeneration. An increase in potential at the suppressor causes an increase in anode current and a decrease in screen current, so that the screen voltage rises. Due to the coupling condenser, the suppressor grid voltage rises still further, and the action is cumulative. Over the region CD, the action is degenerative.

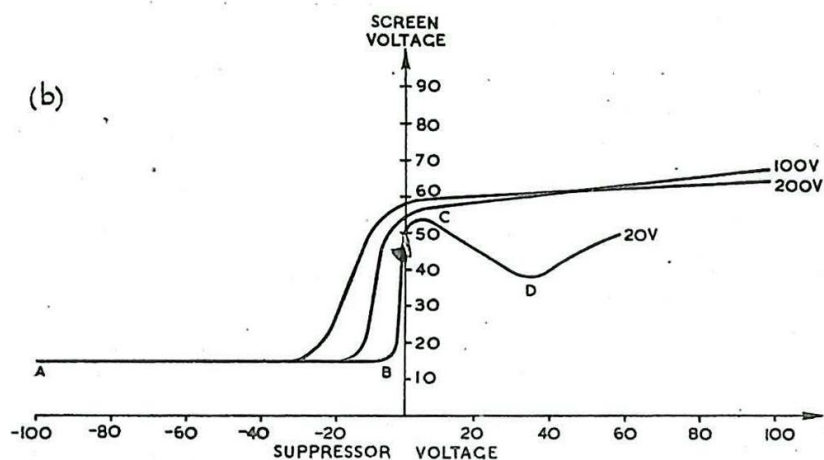
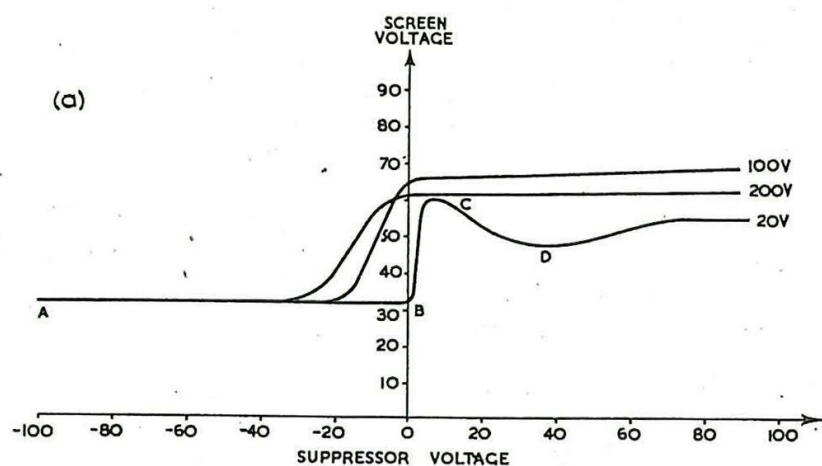


Fig. 300 = Dynamic characteristics for
(a) CV1065 (b) CV1091 for different
values of anode voltage.

Details
of circuits
which
employ this
type of
coupling
are
given in
Chaps. 8
and 10.

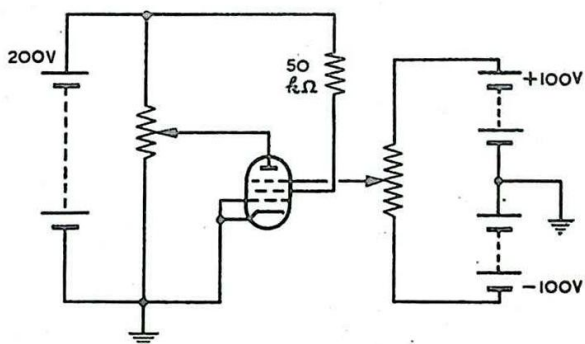


Fig. 301 = Circuit corresponding to
curves of fig. 300.

SOFT VALVES

35. Summary on the effect of gas in valves

If gas at a low pressure is present in a valve it has no appreciable effect on the characteristics so long as the electrons are moving in the valve so slowly that they cannot ionise it. If, however, the anode potential is gradually increased, and the electron velocity thereby increased also, a potential will be reached at which the gas is ionised so that some molecules are split into electrons and (relatively) heavy positive ions.

If the potential difference between anode and cathode is greater than a critical value, called the Ionisation Potential for the gas, the electrons can just acquire the requisite speed, during the intervals between collisions with gas molecules, to maintain the vapour on an ionised state.

As soon as ionisation occurs the space-charge distribution in the valve is profoundly modified. The electrons emitted by bombardment join the stream of primary electrons and travel rapidly towards the anode; the heavier, and therefore more slowly moving, positive ions travel towards the cathode where they neutralise the electron space-charge.

It will be recalled that in a hard diode valve, for small anode potentials, the presence of space charge is the only factor preventing the total emission current reaching the anode. As soon, therefore, as the positive ions neutralise the space-charge the full emission current reaches the anode, although the anode potential may be considerably below that required to produce saturation in the absence of ionisation. The effect is illustrated in the diode characteristics of Fig. 302 in which the curve a, b, c, d, represents the behaviour of a hard hot-cathode valve, and a, b, e, d, that of a similarly constructed valve containing gas at low pressure. For anode potentials less than V_s there is no ionisation and for both valves the current is space-charge limited. For potentials greater than V_s the gas in the soft valve ionises, the space charge is neutralised and the full emission current I_a flows.

Soft valves may have considerable variation in design, but may be classified under three general headings:-

- (i) Cold Cathode Diode Valves
- (ii) Hot Cathode Diode Valves
- (iii) Hot Cathode Triode Valves.

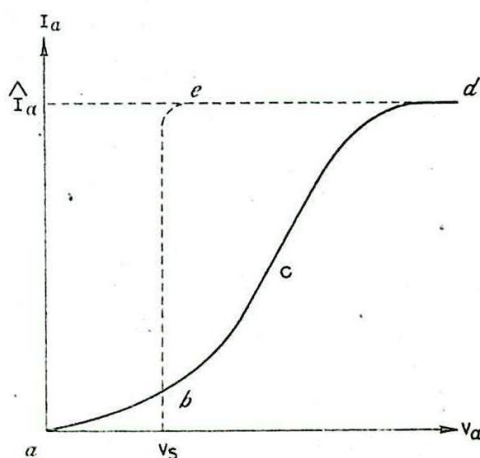


Fig. 302 - Effect of gas on diode characteristic.

Other types of soft valves such as Spark Gaps and Soft Rhumbatrons are of special application in radar systems.

36. Cold Cathode Diode

In this type of valve inert gas is present at low pressure. In the normal state the molecules are electrically neutral and no current flows. As the potential difference between anode and cathode is increased a degree of polarisation of the gas molecules will be set up, and ultimately the Striking Voltage is reached at which ionisation is initiated. This process may be further complicated by extraneous irradiation, such as ultra-violet radiation, and as a result the striking voltage is not in general clearly defined.

The striking voltage, normally about 120 volts, is always greater than the ionisation potential, and corresponds to the "flash-over point" or breakdown of the gas dielectric. Once the valve has struck its subsequent behaviour is similar to that of a hot-cathode diode.

The gas will remain ionised until the current is reduced below a value called the Extinction Current. Provided the anode current is greater than this value the anode-cathode voltage remains constant at approximately the ionisation potential of the gas. This is called the Extinction Voltage.

This valve has use as a HT stabilising device since, once it has struck, its anode-cathode voltage is stabilised at the value of the extinction voltage.

The tube may also be used as the discharge element in simple forms of time-base circuits, but its use in this application is limited by the relatively small separation, about 30 volts, between extinction and striking voltages, and also by the uncertainty in the value of the striking voltage noted above.

Cold cathode diodes are commonly called Neon Valves (tubes). The Stabilivolt, which is discussed in Chap. 19 Sec. 10, is a development of the principle outlined above.

37. Hot Cathode Diode

In general this type of valve contains mercury vapour at a pressure of about 10^{-3} mm. of mercury. A small amount of liquid mercury is present in the tube at room temperature and this is vapourised when the temperature is raised by the application of heater voltage, so that the requisite vapour pressure is produced. This valve is commonly employed for power rectification due to its large current-carrying capacity. Since in this case the onset of ionisation depends on the electron current emitted by the hot cathode the uncertainty in the value of the striking voltage is largely eliminated. The separation of striking and extinction voltages is extremely small. If an attempt is made to raise the voltage above its extinction value while the gas is ionised the cathode will be damaged by positive ion bombardment. For this reason a limiting resistor is normally included in series with the valve to ensure that the voltage across the valve does not exceed a safe value.

The performance of this valve depends very considerably on temperature since this affects the vapour pressure of the mercury and therefore the probability of collisions occurring between electrons and molecules so as to cause ionisation.

38. Hot Cathode Triode

This type of valve is normally termed a Thyratron, although actually this word is a trade name for hot-cathode mercury triodes made by one manufacturer.

A grid is introduced to provide control over the striking voltage; this grid has no control over the extinction voltage. In the absence of ionisation the effect of the grid voltage on the anode current is similar to that for a hard triode valve. For a given grid voltage, negative to cathode, the anode voltage must be raised above a certain value, the cut-off value V_c , before anode current flows. This will not cause ionisation unless V_c is greater than the ionisation potential V_E . Hence, ionisation will not be initiated for a given grid voltage unless the anode voltage is raised above the greater of the two voltages, V_c and V_E . Fig. 303 shows the $I_a - V_a$ characteristic for a typical gas-filled triode at zero bias; i.e., with $V_g = 0$.

Alternatively if the anode voltage be at a value greater than V_E then there is a critical voltage for the grid below which ionisation will not be initiated. This grid control can be made exceedingly sensitive by a suitable design of the electrode assembly.

Once ionisation occurs, variation of the negative potential of the grid has little effect on the space current, since positive ions are attracted to, and surround, the grid so that the electrostatic effect of this electrode is neutralised. The grid therefore loses control until De-ionisation (Recombination) takes place, which can occur only as a result of lowering of the anode potential below the extinction value; in short, after ionisation occurs, the performance is identical with that of a hot cathode diode. The de-ionisation time (10 to 1,000 microseconds) depends on the electrode-potentials and, in general, may be shortened by the application of a large negative potential to the anode.

It will thus be seen that by making the grid voltage sufficiently negative a wide difference between striking and extinction voltage can easily be achieved.

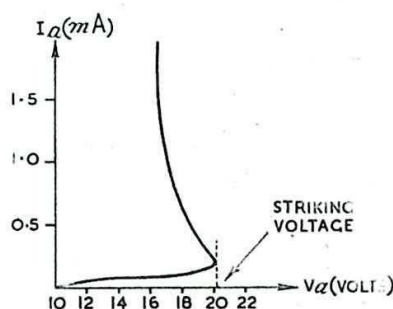


Fig. 303 - Characteristic of typical hot-cathode triode with Argon gas filling, at zero bias.

As mentioned previously, the lower is the anode potential the higher the grid potential needs to be in order that ionisation may occur. The Striking Characteristic of a thyatron is shown in Fig. 304. It is seen that over a large portion of the characteristic the ratio of the increase of anode voltage needed to initiate ionisation to the corresponding change of grid voltage is almost constant. This ratio is known as the Control Ratio and is of the order of 20 to 100.

The grid loses control of the valve current once the arc has started, but its potential is important since this electrode may draw appreciable current. If the grid is at a positive potential it collects electrons and the value of this grid current rises rapidly with grid potential. If, however, the grid is at a negative potential it collects positive ions, and this reverse grid current tends to be small and nearly constant with change of grid potential (Fig. 305). Since large grid-current flow is possible for relatively low values of grid voltage it is advisable to include a limiting resistor in series with the grid so that the grid is maintained at approximately zero potential.

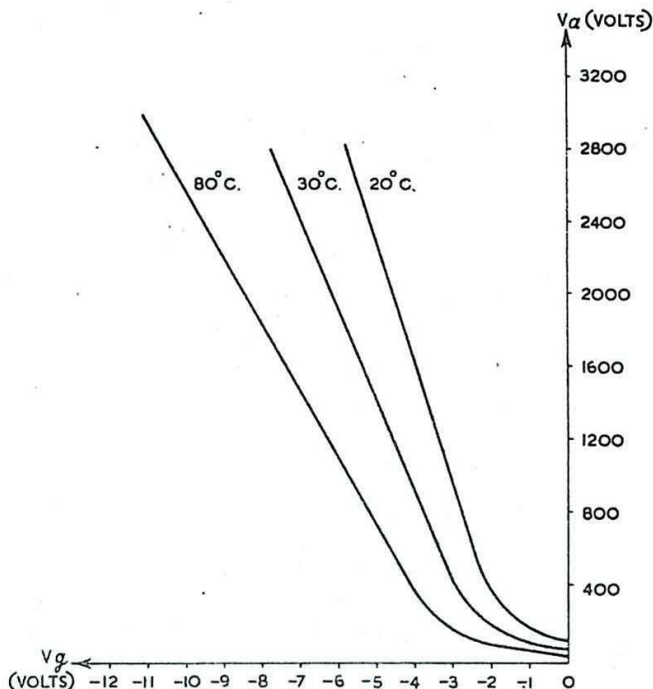


Fig. 304 - Striking characteristics of a typical mercury-filled triode for different temperatures.

Gas triodes are very easily damaged if they are allowed to pass very large currents. If the valve current is allowed to reach the value of the emission current, it means that there is no electronic space charge left at the cathode, and therefore no protection of the cathode surface from positive ion bombardment. This valve current may be limited by inserting a resistor in series with the valve.

The gas used in such triodes is either mercury vapour or one of the inert gases such as neon, argon, helium or hydrogen, or some mixture of these. In the mercury vapour valve, liquid mercury is normally present with the vapour, so that changes in temperature alter the amount of vapour present and hence the operating characteristics of the valve (Fig. 304). In valves filled with neon or one of the other gases the quantity of gas remains constant and the ionisation is therefore hardly affected by changes of temperature, within wide limits. In such valves, however, there is an absorption of gas by the electrodes which causes a change in the gas pressure. This process is significant only over a long period of time.

Because of its adaptability, being usable over a wide range of striking voltages, the hot-cathode triode is employed in many types

of circuits where relatively large currents are required. For very high powers, however, the size of the valve which is necessary and the comparatively elaborate construction have resulted in such valves being superseded by spark gaps. One of the chief disadvantages of large thyratrons is the delay involved in heating the valve to the correct temperature so that the mercury vapour pressure is at the operating value. This usually involves the use of a thermo-stat which incorporates a cooling, as well as a heating device, since overheating may cause erratic operation.

39. Spark Gaps

It is sometimes necessary to arrange a circuit so that a switch is closed when the potential across its terminals reaches a certain high value. This process is conveniently carried out by putting a spark gap between the terminals. As in the case of the cold cathode diode, when the potential reaches the value of the striking voltage the gap conducts, and while it is conducting its resistance is very small. When the potential is removed the spark is extinguished after a period of de-ionisation and a high potential can again be built up before the gap once more conducts. This principle is used in the obsolescent spark telegraphy.

Spark gaps are used in radar for switching (a) in high-power modulators which employ a pulse-forming network and (b) in common T/R systems (where the same aerial is used for both transmission and reception).

In modulating circuits it is usually required that the spark should occur at precisely regular intervals. It is therefore more satisfactory not to control the operation by allowing the voltage across the main electrodes to reach its relatively variable striking voltage, but to initiate the ionisation by the use of a control electrode, whose function is similar to that of the grid of a thyatron. The spark between the main electrodes is struck by the application of pulses of high voltage, but not necessarily of high power, between a third electrode and one of the two main electrodes. Spark gaps which operate on this principle are known as Triggered Spark Gaps, and the third electrode is termed the Trigger Electrode. The triggering pulses, since they need not be of high power, can be readily developed by circuits using normal hard valves.

Alternatively, the distance between the main electrodes can be carried in a cyclical manner so that they approach each other sufficiently closely for the spark to strike at the required instants. Rotary Spark Gaps operate on this principle.

The chief advantages of spark gaps compared with hot cathode valves of comparable dimensions are their robustness and high power-handling capacity. The principal disadvantage of spark gaps is their inflexibility, since for a gap of given dimensions the striking voltage cannot be adjusted. In such a case the design of the modulator is largely determined by the characteristics of the spark gap used.

40. The Triggered Spark Gap

This type of spark gap consists essentially of three electrodes; the two main electrodes take the form of hemispherical caps and the trigger electrode is a wire the end of which protrudes through an aperture in one of the main electrodes. A common construction is illustrated in Fig. 306. Here the main electrode A is connected to a metal cylinder B surrounding the trigger electrode T and separated from it by glass insulation G.

In normal use the spacing between A and C should be such that a spark does not strike between these electrodes, even when the maximum potential difference between them is reached, unless ionisation is initiated by the application of a triggering voltage between A and T. The possibility that sparking may occur between A and C before it is required is reduced by shaping A and C in such a manner that high field concentrations are avoided at the surface of these electrodes. A spark between A and C is extinguished when the voltage developed between these two electrodes falls to a low level, the value of which is determined by the ionisation potential of the gas in the gap.

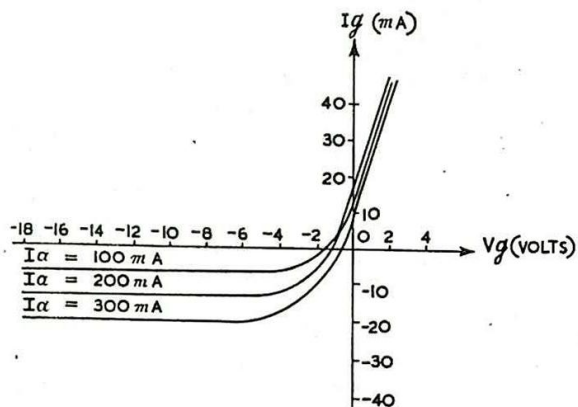


Fig. 305 - Grid current characteristics of a typical thyatron.

The manner in which the triggering voltage has effect in initiating the striking of the spark between A and C is not always clear, but two distinct modes of operation are possible, depending to some extent on the design of the spark gap and to a larger extent on the relative polarities of the applied potentials. It is usual for the electrode A to be earthed, since this permits the circuit producing the trigger pulse to operate near earth potential. The two possible modes of operation are :-

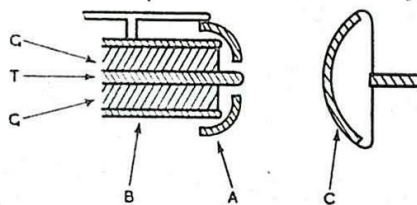


Fig. 306 - Electrode assembly of triggered spark gap (not to scale).

- (i) The triggering voltage applied between A and T causes a breakdown of the small gap between these electrodes and the light emitted by the resultant discharge causes ionisation of the molecules in the main gap AC (possibly there is also some drift of ions into the main gap) and so initiates the main spark. In general this form of discharge is more likely to occur if the potential applied to the trigger electrode is of the same polarity as that of electrode C.
- (ii) If the potential applied to the triggering electrode is of opposite polarity to that of the electrode C, the potential difference between these electrodes is the sum of the magnitudes of their potentials and since in addition there is a high concentration of field and possibly a corona discharge at the point of T, a spark

is likely to strike between T and C. As soon as this occurs a large current flows between T and C causing the potential of T to reach much the same value as that of C, the actual potential acquired by T depending on the relative impedances of the input circuit to T and the output circuit from C. The potential difference between T and A is then normally large enough to break down the gap between these two electrodes. The discharge path between A and C is then completed. It is assumed here that the magnitude of the initial triggering voltage is not large enough to give rise to a discharge between T and A.

The mode of operation (ii) has the advantage that the triggering voltage necessary to initiate the main spark can be smaller in value than that required in case (i). The position of the insulator G, separating T and A has some bearing on the value of the voltage required to trigger spark gaps operated by the mechanism (i). If the insulator is fairly near the front of the electrode A it considerably reduces the necessary triggering voltage, probably due to tracking of the discharge along its surface. Again, for this type of operation, in order to produce efficient irradiation of the main gap by the triggering discharge it is desirable that this takes place between the trigger electrode and the front edge of A. Consequently, the front edge of A should be made sharp.

The action of triggering a three electrode spark gap normally results in jitter, i.e., a random fluctuation in time interval, between the application of a triggering pulse and the instant at which spark discharge occurs. This jitter often may result from random variations in conditions of ionisation in the gap, resulting in a change of the value of impressed voltage necessary to cause the gap to strike. Therefore, since a trigger voltage must take a finite time to build up, jitter will result. Jitter due to this cause may thus be reduced if the triggering voltage can be arranged to build up to the critical value very rapidly, and this can most easily be achieved by providing a triggering voltage much larger than the minimum critical value needed to fire the gap. On the other hand it is clearly desirable from the stand-point of the design of the trigger-producing circuit that the trigger voltage to be produced should be no greater than necessary. In practice, a certain amount of residual ionisation remains between successive arcings and it is necessary to have a higher triggering voltage to break-down the main gap at first than subsequently. If the triggering voltage is made just large enough to trigger off the first spark in the main gap, it is sufficiently above the minimum voltage necessary for triggering subsequent sparks to ensure that only a small amount of jitter occurs.

Since there is asymmetry of the electrodes of the spark gap there is reason to expect that the amount of jitter depends on which of the two main electrodes is used as the anode. Experience shows that in most radar equipments the more reliable operation of the spark gap, from the point of view of accurate timing, occurs if the electrodes A and C serve respectively as anode and cathode of the main discharge.

There is a tendency for the main gap to break down, at voltages lower than expected from the separation of the main electrodes, before the trigger voltage is applied. This effect is probably due to excess of ionisation remaining from a previous spark, and its incidence limits the recurrence frequency. The possibility of this premature sparking is much less if the value or duration of the spark current is reduced, and if it is possible to remove the residual

products of the previous spark by, say, blowing air into the gap. This of course precludes the possibility of enclosing the spark-gap structure. De-ionisation time is also affected by the type of gas in which the spark is produced.

Triggered spark gaps operate very well in air at atmospheric pressure provided they are not enclosed. If they are enclosed, their life may be fairly short. This is probably due to the production of nitrogen peroxide by the discharge. The presence of this gas in the gap increases its sparking potential and makes it inoperative at its previous potential ratings.

Although it is generally not necessary for a spark gap to be enclosed in an airtight envelope if it is always to be used on the ground, for airborne use such an envelope (usually glass) is desirable, due to the variation of pressure with height.

It appears necessary, in order that the de-ionisation time be sufficiently rapid, for the gas in which the spark is struck to contain a certain amount of oxygen, which has the property of promoting rapid de-ionisation. Various mixtures of gases are suitable, two of which are: one, a mixture containing 90% nitrogen and 10% oxygen at atmospheric pressure, and the other a mixture of 97% argon and 3% oxygen at three atmospheres pressure.

The life of an enclosed gap depends largely on the rate at which the oxygen in the gap is used up either by oxidation of the electrodes or by combination with the other gases present in the case of the air-filled gap. Also the electrodes are liable to wear both by oxidation and by sputtering in the spark. In order to reduce the wear they are usually made of molybdenum or tungsten.

41. The Rotary Spark Gap

In this form of spark gap the electrodes are generally fixed, and a metallic arm is rotated between them at a uniform speed by means of an electric motor. The axis of rotation of the moving arm cuts the line joining the electrodes mid-way between them, so that in one position the two ends of this arm just clear the respective electrodes simultaneously (Fig. 307). When the rotating arm reaches some such position as that shown in the figure then a spark

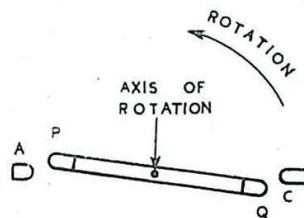


Fig. 307 - Electrode assembly of rotary spark gap.

strikes between A and P and between C and Q, forming a conducting path from A to C; i.e., the switch whose terminals are A and C is closed. The spark is extinguished when the voltage driving current through the switch falls to a low value, and not, in general, because the movement of the metallic arm has caused an increase in the spacings AP and CQ. In normal operation the required duration of the spark is of the order of 1 microsecond, in which time the metallic arm does not move appreciably.

The recurrence frequency of the spark is twice the frequency of rotation if a single arm is used. However, this frequency is increased if there is more than one arm. Thus, with a rotating spark gap of the form shown in Fig. 308 the recurrence frequency of the

spark is four times the frequency of rotation of the arms, so that if the motor driving the rotating electrode has a rating of 6000 r.p.m., the spark recurrence frequency is 400 per second. Recurrence rates as high as 2500 per second have been obtained.

If there is more than one pair of fixed electrodes arranged round the circumference of a circle, and more than one arm, it is possible to provide a number of spark-gap switches whose times of closing are synchronised. An arrangement such as this is required for the Marx circuit, (Chap. 14 Sec. 38).

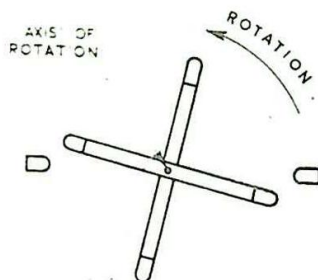


Fig. 308 - Double arm rotary spark gap.

The closeness with which two electrodes have to approach before a spark strikes, if there is a potential difference between them, is rather erratic, and varies slightly in this form of a spark gap for successive strikings of the spark. Consequently, there is a slight variation in the time interval between successive sparks. This jitter is smaller the larger is the radius from the axis of rotation to the point at which the spark is struck, since then the circumference covered by the moving arm between successive strikings of the spark is large compared with the variations in the sparking distance. Furthermore, if the repetition rate of the spark is increased by increasing the number of arms mounted on the same shaft, the length of the moving electrode must be proportionately increased if the relative amount of jitter is not to increase.

Owing to the difficulty of reducing jitter beyond a certain point with reasonable dimensions between the fixed electrodes, the applications in which a rotary spark gap can be used are rather more restricted than those in which a triggered spark gap can operate. For example, since the time of discharge of the spark is liable to some variation as the rotating arms approach the fixed electrodes, it cannot readily be co-ordinated with an independent source of timing. In fact when a rotary spark gap is used in radar equipment the timing of the spark itself must be used as the basis of synchronisation of the whole system if accurate measurement of time-intervals is required. However, the random nature of the timing of a rotary spark gap in a radar equipment does decrease the chance of periodic external interfering signals being "locked" to the transmitted pulse and so giving the appearance of an Echo (re-radiated signal).

When the spark is struck the current flowing between the electrodes is usually very heavy (50 amperes or more), and consequently the electrodes of the spark gap tend to wear. To minimise this it is usual for the ends of the electrodes to be made of molybdenum. The life of a rotary spark gap is determined almost entirely by the wear of the molybdenum electrodes. This wear is governed chiefly by the value and total duration of the current so that the life of the device is roughly inversely proportional to the peak current, duration and repetition rate of the sparks.

The factor which ultimately limits the maximum recurrence frequency at which a spark gap can operate is the time taken for the

gases to de-ionise. However, in a rotary spark gap the sparks are unlikely to be maintained except during the short interval when the rotating arm is close to the stationary electrodes. Besides, the spark normally takes place in air and is not enclosed, so that the motion of the moving arm creates a draught which tends to clear ions away from the fixed electrodes after a discharge has ceased in readiness for the next spark. For these reasons the spark gap can be operated at a high spark recurrence frequency (e.g. 2,500 per second).

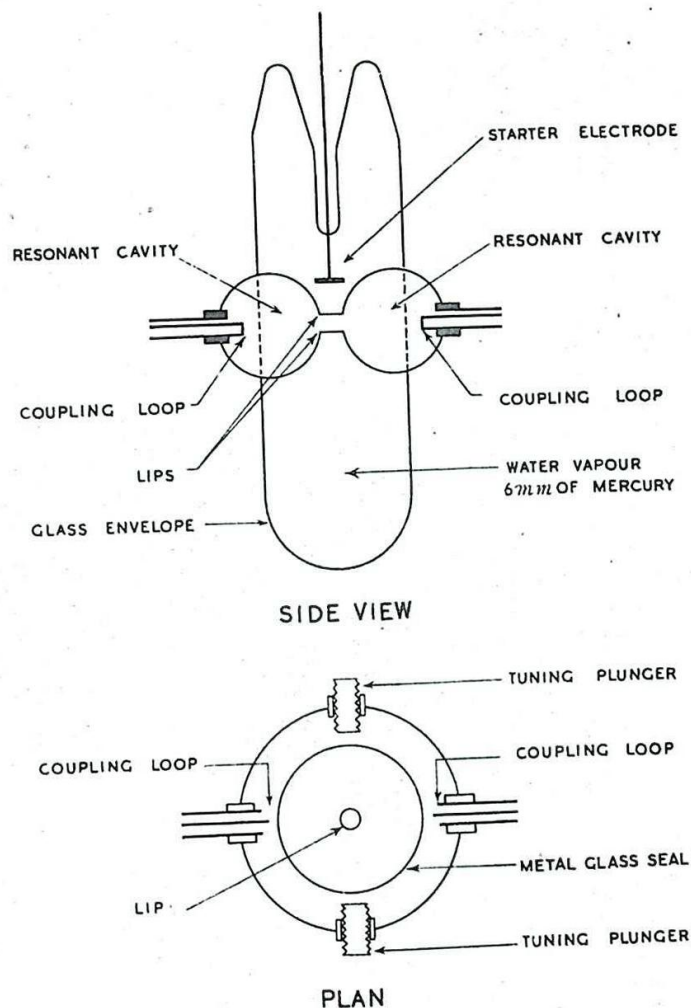


Fig. 309 - Soft Rhumbatron.

42. Soft Rhumbatron

In many radar equipments one aerial system is used for both transmission and reception of signals. This arrangement involves the use of a switch which is so arranged that during transmission of RF pulses the transmitted power is prevented from reaching the receiver, whereas during reception the received power is allowed to pass freely to the receiver. If the transmitted power were permitted to reach the receiver, designed to deal with extremely low level signals (not greater than $1/10$ watt), it would in general damage the receiver, or at least render it inoperative for the reception of signals.

The soft rhumbatron is a type of electronic switch particularly suited to common T/R working at frequencies of 3000 Mc/s. or more.

Transmitted peak powers are normally considerably greater than $1/10$ watt (they may be 50 kW. or more) so that the switch between the transmitter and receiver must operate so as to prevent the major portion of the transmitted peak power from reaching the receiver. Furthermore, the switch must be capable of passing received signals, which are in general of considerably less power level than $1/10$ watt, and should be able to do so within a fraction of a microsecond after the transmitter has ceased to operate.

The construction of a typical soft rhumbatron switch is shown in Fig. 309. It consists essentially of a rhumbatron, (cavity resonator, Chap. 5 Sec. 25) the lips of which are enclosed in a glass envelope containing gas at low pressure. The normal filling of the envelope is water vapour at a pressure of 6 mm. of mercury. Coupling of RF power in and out of the resonator may be made by means of small loops, which are fed from coaxial lines, or by waveguide feed, while the tuning of the resonator may be accomplished by means of screw plungers.

If the rhumbatron switch is inserted in a coaxial line fed by a RF voltage generator, the following behaviour is observed. For low RF powers, less than $1/10$ watt, if the resonator is tuned to the appropriate frequency and the coupling loops correctly adjusted the switch acts almost as a one-to-one transformer, and the transmission is nearly perfect; (Fig. 310). In practice there is a small loss of power known as reception or low-level loss of the order of 0.5 db. in passing through the rhumbatron. If the input voltage is increased, the RF voltage developed across the lips of the resonator becomes large

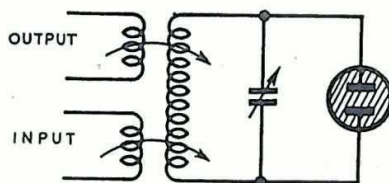


Fig. 310 - Equivalent of soft rhumbatron circuit.

enough ultimately to initiate glow discharge. Once ionisation has taken place this voltage remains approximately constant at a low value, being analogous to the extinction voltage of simple gas valves. The power transmitted through the device is then a function of the running voltage, and since this voltage is effectively constant, the power transmitted through the switch is small and practically independent of the input voltage. The glow discharge can be considered as placing a low shunting resistance directly across the cavity. (Fig. 310).

The protection afforded to the receiver by the soft rhumbatron switch during the period of operation of the transmitter of a radar equipment is determined by the running voltage of the glow discharge and the break-down time of the gap. The running voltage is closely related to the pressure of the water vapour used in filling the switch, whilst the breakdown time, besides being dependent on conditions of operation (see below), is also related to the effective Q (normally about 300), of the resonator. The Q of the resonator determines the rate of rise of voltage across the lips of the resonator. The break down time should be as short as possible since, until a glow discharge is formed, transmission through the switch is unimpaired. In general the protection improves as the effective Q is increased or as the water vapour pressure is lowered from 15 to 2 mm. of mercury; (see Fig. 311).

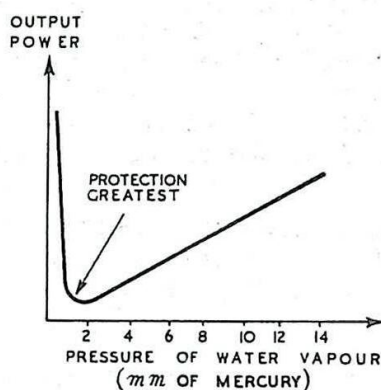


Fig. 311 - Effect on soft rhumbatron of variation of gas pressure.

The recovery time of the soft rhumbatron switch is also found to be dependent on the pressure of the water vapour filling and on the effective Q of the resonator. In general, the recovery time becomes worse as the effective Q is increased or as the water vapour pressure is reduced from 15 to 2 mm. of mercury.

As the effective Q is increased the resonator tuning becomes more critical.

A water vapour filling is found to give a short recovery time (Fig. 312). Other gases which are suitable from this point of view are found to be unstable when subjected to RF discharge; i.e., there is a change in performance of the switch with time. When water vapour is used it may be necessary to provide heating for the switch to prevent the temperature of the water vapour from falling to too low a value (below 0°C).

The normal gas filling for a soft rhumbatron switch having a resonator with an effective Q of 300 and used in connection, say, with power of frequency about 3000 Mc/s., is water vapour at a pressure of 6 mm. of mercury. This arrangement is regarded as a reasonable compromise between protection, recovery time and criticalness of tuning adjustment. It has already been stated that the protection offered by the soft rhumbatron switch depends on the break-down time of its gap. Provided ions are already present between the lips of the switch this break-down time is very small and the protection is good.

The soft rhumbatron switch, as considered so far, functions well for repetition frequencies from about 500 per second upwards, but does not protect the receiver at the time when the transmitter is first switched on. In other words, once operation has started ions persist in the space between the lips of the resonator during the intervals between transmitter pulses, provided these intervals are not longer than about 2000 microsecs; but at the start of operation there are few if any ions present.

In order to provide the necessary ions at the start of operation, it is usual to run a steady current discharge between an auxiliary electrode (Starter or Keep-Alive Electrode) and the wall of the resonator. The current of this priming discharge is normally about 1. to 2 mA. and operates at a running voltage of about 500 volts between the auxiliary electrode and the resonator. This priming discharge must of course be struck before the transmitter is switched on. There is no disadvantage if the priming discharge is left operating continuously while the radar equipment is in use, since the presence of ions in the gap does not have any appreciable effect on the transmission of received signals (of power less than 1/10 watt) through the switch.

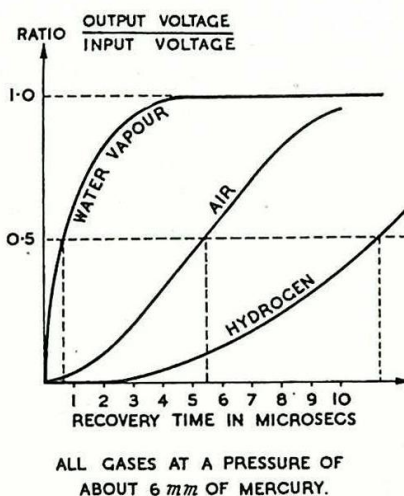


Fig. 312 - Comparison of recovering time for various gasses.

NON-OHMIC RESISTORS

43. General

The resistors used in most electrical circuits, of which radar circuits are no exception, are made of conducting or quasi-conducting materials which conform more or less faithfully to Ohm's law. In some instances, particularly at high frequencies, the self-inductance or earth-capacitance of a resistor is not negligible, and must be allowed for as indicated in the equivalent circuit of Fig. 313. There are, however, instances in which the departure of a material from the simple proportionality law of ordinary resistors is such as to make these substances particularly useful in certain applications. The use of copper-oxide or "cat's whisker and crystal" rectifiers is well known and needs no description here. In both cases the surface junction of metal with non-metal introduces the non-linearity which makes the combination useful in rectifying or frequency-changing circuits. Electrons leave the metal more readily than the non-metal, so that the forward and reverse resistances have widely different values.

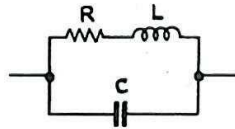


Fig. 313 - Equivalent circuit of a practical resistor.

Recently various materials have been developed which show marked non-linearity in their voltage-current characteristics without the discontinuity associated with the former combinations. The resistance decreases with an increase in current due to the unusual behaviour of the conduction electrons in the atomic structure. In addition, the negative temperature coefficients commonly associated with quasi-conductors may have a marked effect on the resistance-current characteristic, since an increase in current raises the temperature and so reduces the resistance.

In the case of Thermistors, described below, this leads to an actual fall in voltage as the current rises, so that beyond a certain point the material exhibits the property of negative resistance.

44. Metrosil

This is the trade name of a quasi-conducting material originally developed as a lightning arrester and now widely used as a protective device in power supply circuits. The V-I static characteristic for metrosil is shown in Fig. 314. The curve follows a power law $V \propto I^\beta$, so that the R-I curve also follows a power law. The particular value of β for a given metrosil depends on the constitution of the material used. By suitable manufacturing methods β can be given any value between 0.2 and 0.4.

The constant of proportionality is given by the equation :-

$$V = R_1 I^\beta$$

and a metrosil can be made to give any required value of R_1 . This

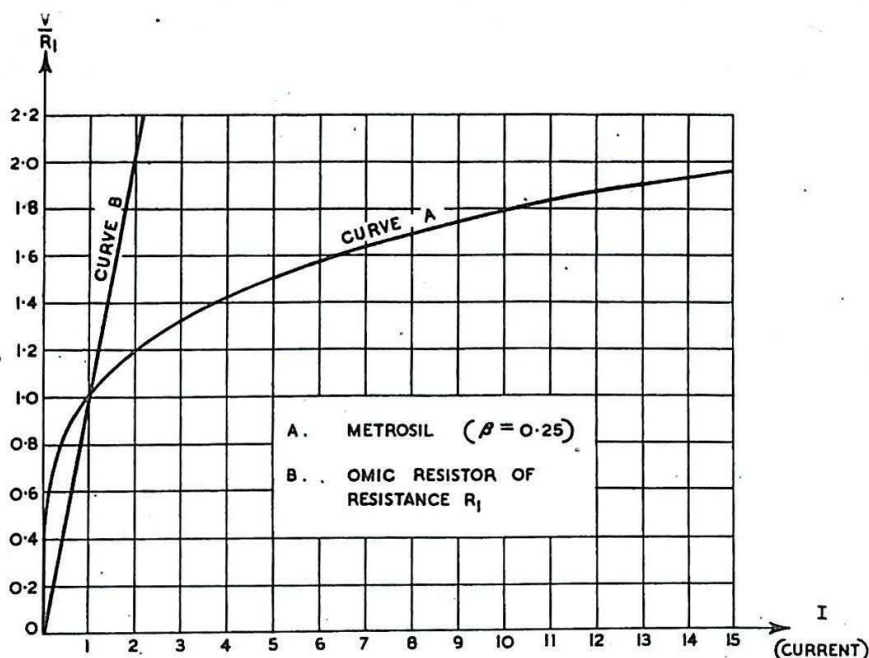


Fig. 314. - V-I characteristics, (a) metrosil: ($\beta = 0.25$) (b) ohmic resistor of resistance R_2 .

depends, as in any other resistor, on the shape and size of the element. If R is the resistance of the element for a voltage V and a current I , then since $V = RI$ it follows that

$$R = R_1 I^{(\beta-1)},$$

and by substitution, we obtain

$$R = \frac{R_1^{1/\beta}}{\left(\frac{I}{\beta} - 1\right)}.$$

R_1 is the resistance presented by the metrosil to a current of one amp.

Fig. 315 represents the resistance-voltage characteristic for a one-ohm metrosil with $\beta = 0.25$. For any other value of R_1 (but the same β) the coordinates of Fig. 315 should be multiplied by R_1 . Thus, a 100 ohm metrosil would have a resistance of 100 ohms for an applied potential difference of 100 V., but only 12.5 ohms if the applied voltage were increased to 200V.

The response of a metrosil to alternating voltages is instantaneous and involves no phase change, so that the static characteristics are usable for AC as well as for DC. If a sinusoidal voltage is applied to a metrosil the resulting current is very much distorted, and the harmonic content must be taken into consideration when calculating the RMS value of the current. The RMS value of the current (I_{RMS}) is obtained by multiplying by a factor k_I the current

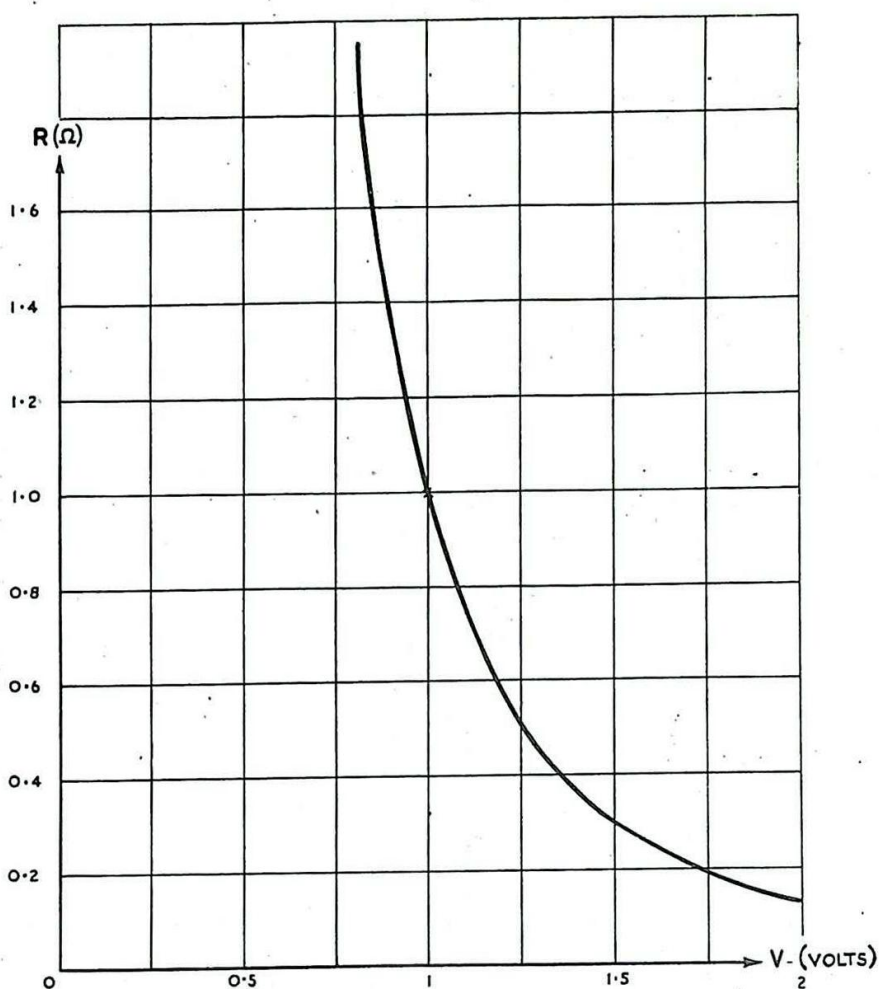


Fig. 315 - Resistance voltage characteristic for a one-ohm metrosil with $\beta = 0.25$.

which would flow in response to a constant voltage equal to the RMS value of the applied voltage. The variation of k_I with β is shown in Fig. 316.

The power dissipated for a sinusoidal applied voltage is given by

$$P = k_P V_{\text{RMS}} I_{\text{RMS}},$$

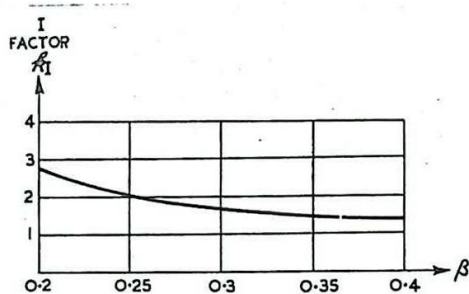


Fig. 316 - Variation of k_I with β .

where k_p depends solely on β and varies as shown in Fig. 317. (k_p is the Power Factor)

for the metrosil, if this term is used in its general sense).

A metrosil usually takes the form of a ceramic compound moulded into a disc or annular ring. Dampness can be excluded by impregnation or by immersion in oil. Heat has little effect, the compound being stable and mechanically strong

up to red heat. The power rating depends to a large extent on the mounting. With free air circulation it is generally taken to be 1 watt/square in. of surface.

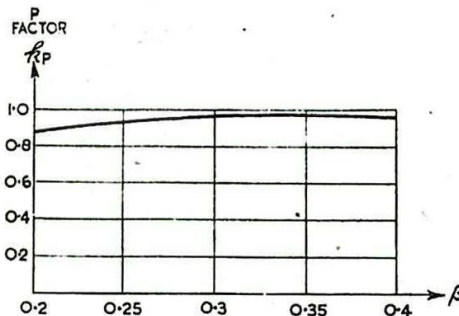


Fig. 317 - Variation of k_p with β .

When short duration overloads are to be withstood, the energy can be calculated from the approximate equation. (Energy in watt-secs) = $712 \times (\text{Weight in } k_g) \times (\text{Admissible temperature rise in } ^\circ \text{C.})$.

It is found that with a rise of temperature from T_0 to T_1 $^\circ \text{C.}$ R_1 changes approximately according to the equation,

$$(R_1)_{T_1} = (R_1)_{T_0} [1 - 0.0018 (T_1 - T_0)].$$

Various applications of metrosils are as follows :-

When used as a voltage limiter :-

- (i) Lightning arrestor.
- (ii) Safety device to suppress transient surges in inductive circuits.
- (iii) Field discharge resistors.
- (iv) Protection of contactor coils.

When used as a variable resistor :-

- (v) Overload protection.
- (vi) Voltage regulation.
- (vii) Production of harmonics.
- (viii) Pulse shaping.
- (ix) Bridged - contact switching.

45. Thermistors

Thermistor is a contraction of the words "Thermal Resistor", and is the name given to certain quasi-conducting materials with large negative temperature coefficients. As the current through such a resistor increases the temperature rises and the resistance diminishes. A point is reached where the voltage is a maximum, and beyond this point further increase in current necessitates a fall in voltage. For current fluctuations about a mean greater than this critical value the differential or slope resistance, $\frac{\delta V}{\delta I}$, is negative.

Heat may also be applied to the thermistor externally, so that the resistance then depends chiefly on the controlled temperature and not on the heat generated by the current. This indirect method of heating is utilised in thermostatic or temperature-measuring devices. It is with the properties associated with direct heating that we are concerned here.

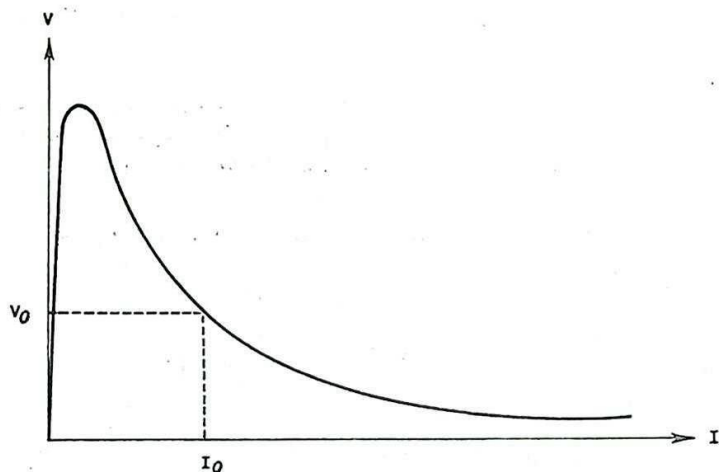


Fig. 318 - Typical V-I characteristic for a thermistor.

Fig. 318 shows a typical static voltage-current characteristic for a thermistor. The rise in temperature throughout the material is subject to a time lag T_d : for alternating voltages this produces an increasing phase delay as the frequency f rises from zero. This is illustrated in Figs. 319 and 320 for a small oscillatory current about a mean value I_0 somewhere on the negative-resistance portion of the curve of Fig. 318. As f increases, $\frac{1}{f}$ becomes smaller compared

with T_d , and the phase lag between current and voltage increases.

This is further illustrated in Fig. 321. A portion of the static curve (a) holds for direct currents. As the frequency rises the graph assumes the form (b). When the phase delay reaches 90° the impedance is almost a pure reactance (c) (save for the inherent non-linearity of the material). Further increases in frequency are accomplished by a disappearance of the negative-resistance property (d) and ultimately the thermistor behaves like an ordinary ohmic resistor (e).

The frequency at which the resistance to alternating currents changes from negative to positive is of the order of a few kc/s.

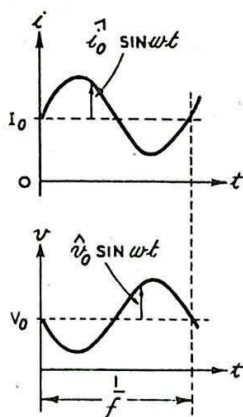


Fig. 319 - Graph of v against i when $T_d \ll \frac{1}{f}$.

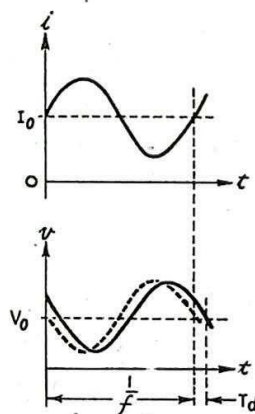


Fig. 320 - Graph of v against i when T_d is not negligible.

Thermistors usable up to 25 kc/s. have been manufactured, but at such a frequency they are unstable.

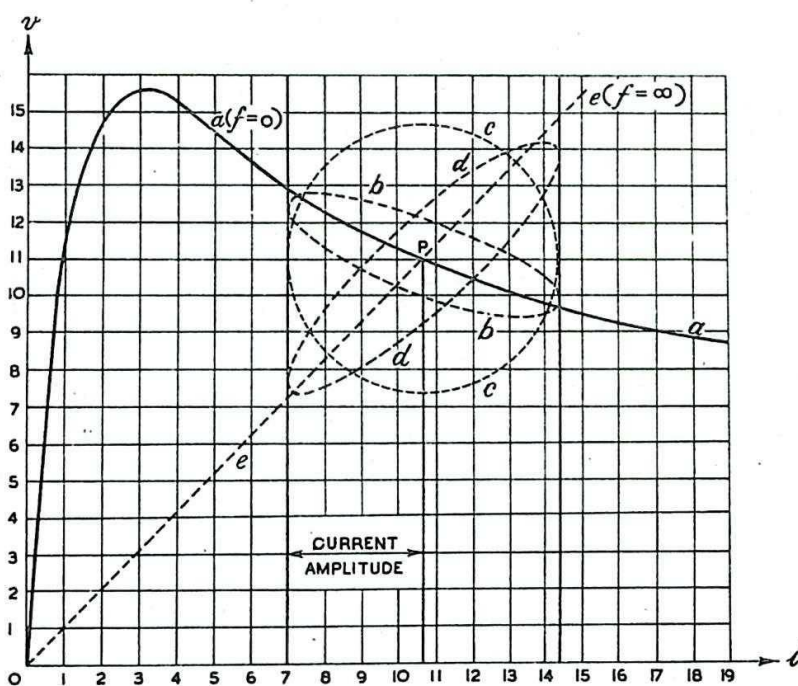


Fig. 321 - Static characteristic (full) and dynamic characteristics (dotted) of thermistor for different frequencies.

Further limitations on the use of thermistors are imposed by their dependence on the ambient temperature. In all accurate applications except those of temperature measurement and control, some form of compensation must be provided. This may be accomplished by a combination of two or more thermistors with different temperature characteristics.

CHAPTER 7

VALVE AMPLIFIERS

(In this chapter many of the diagrams are drawn with the screen and suppressor grid connections omitted for simplicity. When this occurs it may be assumed that normal supply connections are made, the suppressor grid being connected to cathode and the screen to a dropping resistor or potentiometer with bypass condenser).

1. INTRODUCTION

The types of voltage variations most frequently requiring amplification in radar are illustrated in Figs. 322 and 323. The variations shown in Fig. 322 (a) and (b) are commonly termed voltage (or current) pulses; those illustrated in Fig. 323 (a) are called

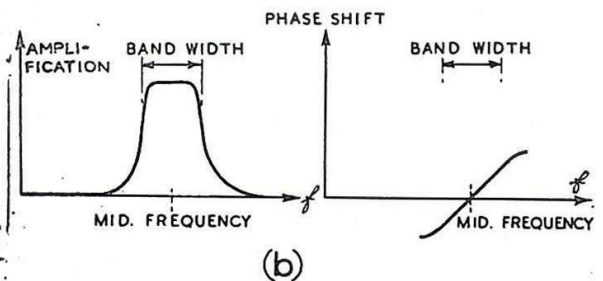
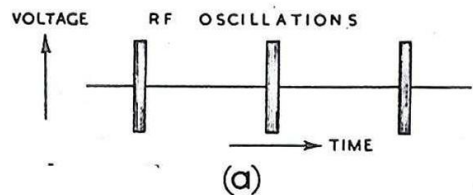
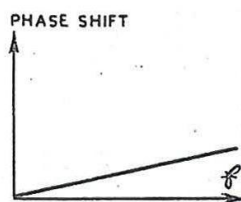
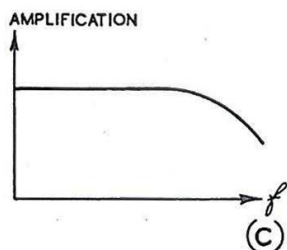
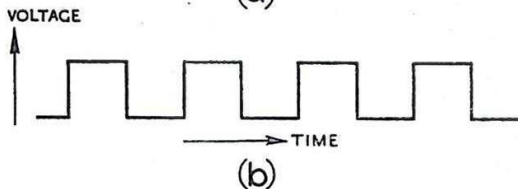
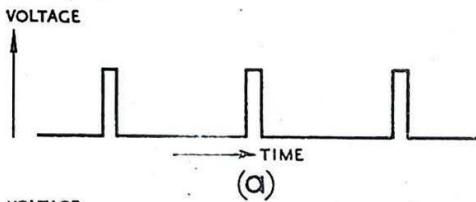


Fig. 322 - Typical voltage variations and the amplifier characteristics necessary for their amplification.

Fig. 323 - R F Pulses and the amplifier characteristics necessary for their amplification.

RF pulses. Either of these types of voltage variation may be analysed into sinusoidal components covering a range of frequencies (Fourier analysis, Chap. 16 Sec. 1). In considering the behaviour of amplifiers it is sometimes desirable to consider the response of the circuit to these sinusoidal components. Alternatively, in the case of the pulses of the type shown in Fig. 322 it may be preferable to deal with the transient response of the amplifier to an idealised pulse without resorting to a detailed analysis.

Considering both pulses and RF pulses from the aspect of their sinusoidal components, the amplifiers required to handle such variations must possess characteristics of the types shown in Figs. 322 (c) and 323 (b) respectively.

2. TYPES OF DISTORTION

An amplifier is liable to produce three types of distortion, any one of which causes the form of the output voltage to be different from that of the input.

(i) Amplitude Distortion

This type of distortion is due to non-linearity of the valve characteristics over the operating range of grid and anode voltages. The effect is to introduce frequencies into the output which were not present in the input.

(ii) Frequency Distortion

This type of distortion is caused by unequal amplification of the components of different frequencies present in the input voltage.

(iii) Phase Distortion

This type of distortion occurs when the relative phases of the input components of different frequencies are not preserved in the output. However, if the phase-shift produced by the circuit is proportional to frequency all the components are delayed equally in time and the output is delayed on the input, but there is no other distortion.

3. CONVENTIONAL SYMBOLS AND EQUIVALENT CIRCUITS

Fig. 324 (a) shows the circuit arrangement of an amplifier with load z_l .

The following symbols are used :-

v_a is the potential difference between anode A and cathode K, and is reckoned positive if the anode is at the higher potential.

V_a is the steady component of v_a .

v_g is the potential difference between control grid G and cathode, and is reckoned positive if the grid is at the higher potential.

V_g is the steady (Bias) component of v_g . (V_g is usually negative).

V_B is the HT supply voltage.

i_a is usually (unless otherwise stated) taken to indicate the

fluctuating component of the anode current, and is considered positive in the direction of conventional current flow, from anode to cathode.

I_a is the steady component of anode current.

v_i is the change of voltage at the grid (input voltage); i.e. the amount by which v_g exceeds V_g .

v_o is the change of voltage at the anode (output voltage); i.e., the amount by which v_a exceeds V_a .

v_s is the signal voltage applied to the input circuit (when different from v_g or v_i).

Capital letters A, G, K, used as subscripts, denote that the voltage concerned is measured from earth (or chassis) and not necessarily from cathode.

μ is the Amplification Factor of the valve, defined as the ratio of a small change in anode voltage to the corresponding small change in grid voltage provided the anode current in the valve remains unchanged.

R_a is the Slope Resistance (Differential or Output Resistance) of the valve, defined as the ratio of a small change in anode voltage to the corresponding small change in anode current provided the grid voltage of the valve remains constant.

G_m is the Mutual Conductance (Transconductance) of the valve, defined as the ratio of a small change of anode current to a small change of grid voltage, the anode voltage remaining constant.

If i_a , v_i and v_o are corresponding small changes in anode current, grid voltage and anode voltage respectively, the valve constants may be written :-

$G_m = \frac{i_a}{v_i}$, where
anode voltage is constant,

$\mu = -\frac{v_o}{v_i}$, where
anode current is constant,

and
 $R_a = \frac{v_o}{i_a}$, where
grid voltage is constant.

It can be shown that

$$\mu = g_m R_a$$

*** Proof of formula $\mu = G_m R_a$

For any valve there is a relation between the anode current i_a , anode-cathode voltage v_a , and the grid-cathode voltage v_g .

We may denote this relation by the equation

$$f(i_a, v_a, v_g) = 0$$

.....(1)

From the theory of partial derivatives we obtain

$$\frac{\partial f}{\partial i_a} di_a + \frac{\partial f}{\partial v_a} dv_a + \frac{\partial f}{\partial v_g} dv_g = 0$$

..... (2)

Putting $di_a = 0$, $dv_a = 0$, $dv_g = 0$ in succession in (2), we have

$$-\mu = \frac{\partial v_a}{\partial v_g} = -\frac{\partial f / \partial v_g}{\partial f / \partial v_a}$$

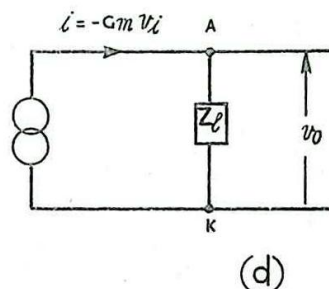
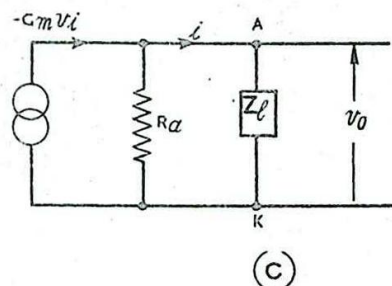
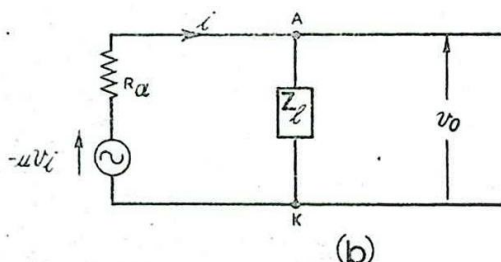
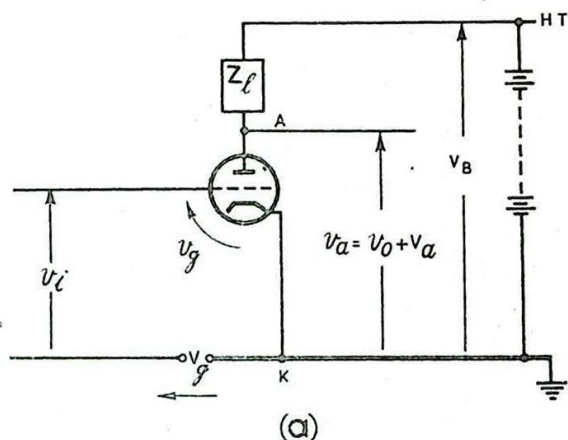


Fig. 324 - Equivalent circuits.

$$\text{so that } \mu = \frac{\partial f}{\partial v_g} / \frac{\partial f}{\partial v_a} ;$$

$$G_m = \frac{\partial i_a}{\partial v_g} = - \frac{\partial f}{\partial v_g} / \frac{\partial f}{\partial i_a} ,$$

$$\text{and } R_a = \frac{\partial v_a}{\partial i_a} = - \frac{\partial f}{\partial i_a} / \frac{\partial f}{\partial v_a} .$$

$$\text{Hence } G_m R_a = \frac{-\frac{\partial f}{\partial v_g}}{\frac{\partial f}{\partial i_a}} \cdot \frac{-\frac{\partial f}{\partial i_a}}{\frac{\partial f}{\partial v_a}} = \frac{\frac{\partial f}{\partial v_g}}{\frac{\partial f}{\partial v_a}} = \mu .$$

Stage Amplification and Equivalent Circuits

The expression for the voltage amplification provided by the stage is :-

$$m = \frac{v_o}{v_i}$$

If v_i , and hence v_o , vary sinusoidally then :-

$$m = - \frac{\mu Z_f}{R_a + Z_f} \dots\dots\dots (3)$$

In general m is complex; its magnitude will be denoted by $|m|$, and is equal to $\frac{\hat{v}_o}{\hat{v}_i}$.

The circuit of Fig. 324 (a) may be represented, from the point of view of fluctuating voltages, by the equivalent circuit shown in Fig. 324 (b). In this circuit the current i is given by

$$i = - \frac{\mu v_i}{Z_f + R_a} ;$$

$$\text{hence } \frac{v_o}{v_i} = \frac{i Z_f}{v_i} = - \frac{\mu Z_f}{Z_f + R_a} = m, \text{ justifying the representation.}$$

Equation (3) can be rearranged as follows :-

$$\begin{aligned} m &= \frac{-\frac{\mu}{R_a} \cdot Z_f \cdot R_a}{R_a + Z_f} \\ &= - G_m \frac{R_a Z_f}{R_a + Z_f} \dots\dots\dots (4) \end{aligned}$$

Equation (4) suggests the form of equivalent circuit shown in Fig. 324 (c). This is normally referred to as the Constant-Current Generator Equivalent Circuit whilst the other (Fig. 324 (b)) is termed the Constant-Voltage Generator Equivalent Circuit. The two circuits express the same relationship between v_i and v_o .

If the valve used is a pentode whose slope resistance is large compared with the load impedance the arrangement shown in Fig. 324 (c) simplifies to that shown in Fig. 324 (d).

If the simple equivalent circuits of Fig. 324 (b), (c) and (d) are compared with the actual circuit (a) it is seen that z_l lies between A and K in the equivalent circuits and not A and HT+. This representation is possible because the impedance of the HT supply to alternating voltages is very low, so that, as far as such voltages are concerned, the upper end of the load is connected to cathode.

The fluctuating component of the anode current (Fig. 324(a)) is $-i$, i having the sign convention as shown in the figure.

4. BIAS FOR AMPLIFIERS

In using a valve as an amplifier, or for other purposes, it is common practice to use the current flowing through the valve, in conjunction with suitable circuits, to maintain the cathode at a fixed steady potential relative to the mean potential of the grid. The steady potential of the grid in the absence of any input voltage is usually earth potential, since the grid is normally connected to earth by means of a grid-leak resistor R_g and isolated from steady potentials on previous circuits by means of a condenser C: (Fig. 325). If a succession of rectangular voltage pulses is applied to C and R_g in series, the time-constant CR_g being of suitable duration, a succession of rectangular pulses is developed at the grid, having the same mean potential as the steady potential at the grid in the absence of the pulses. Assume that the succession of pulses at the grid of the amplifier is as shown in Fig. 326. The duration of each positive-going pulse is T_1 and the repetition period of the pulses is T_2 ,

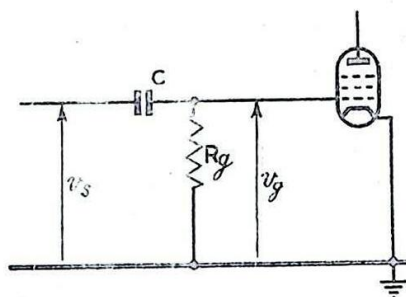


Fig. 325 - Common input network.

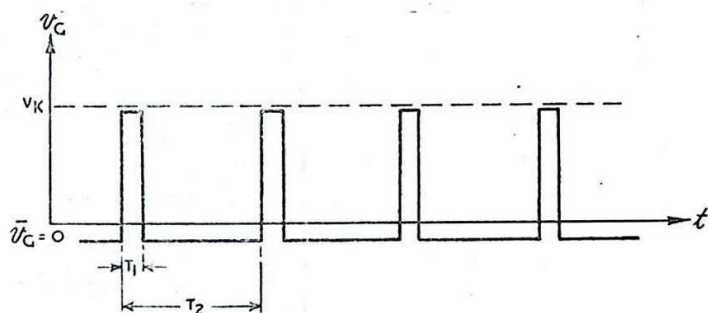


Fig. 326 - Choice of Bias (Input between grid and earth).

T_2 being much larger than T_1 . Then if the cathode of the valve is at earth potential each positive-going pulse of voltage will carry the grid potential above that of the cathode and grid current will flow. This will introduce considerable distortion if the output impedance of the previous stage is at all large (see Chap. 9 Sec. 3).

In this case, therefore, it is necessary to make the steady potential of the grid more negative than the steady cathode potential i.e. a negative bias is required. The steady value of the cathode voltage should be chosen as illustrated in Fig. 326. If the succession

of rectangular pulses is inverse to that shown in Fig. 326, i.e., the pulses are negative-going, it is not necessary, and may be inadvisable, for the valve to have an appreciable negative bias. In this case a negative bias might lead to the negative-going pulses cutting off the valve current.

If a negative bias is required it may be obtained by allowing the current flowing through the valve to pass through a resistor connected between the cathode and earth; (Fig. 327). The current in the valve varies with the potential at the grid, and unless negative feed-back for alternating currents is required (Sec. 16) it is necessary to have a decoupling (or bypass) condenser connected across the bias resistor; (Fig. 327). The condenser C_K charges so that the potential difference between its plates is $R_K \bar{i}$, where \bar{i} is the mean current flowing through the bias resistor R_K . If the time-constant $C_K R_K \gg T_2$ the condenser cannot discharge appreciably within the repetition period T_2 of the applied pulses, and the cathode of the amplifier is maintained at a steady potential.

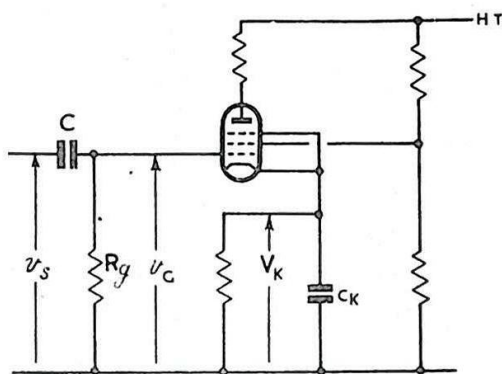


Fig. 327 - Use of cathode self-biasing circuit.

The method of using a cathode resistor in order to obtain negative bias has its limitations. Provided the valve is operated over the linear portion of its dynamic characteristic, variation of signal amplitude does not affect the mean valve current and the bias therefore remains constant. If the input voltage is allowed to vary over the curved portion of the characteristic, the mean current and hence the bias are affected. This means that the bias varies with signal amplitude. For Class B or C operation, therefore, some other system of biasing should be employed.

A convenient method of biasing a valve beyond cut-off or on the curved portion of the dynamic characteristic is illustrated in Fig. 328. Provided R_1 is small enough the variations of current through the valve are small compared with the current through R_1 and the bias remains approximately constant. The condenser C_K smooths out those fluctuations which still occur.

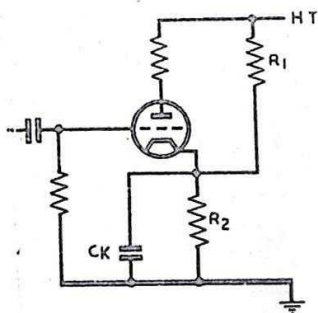


Fig. 328 - Circuit for class B or C biasing.

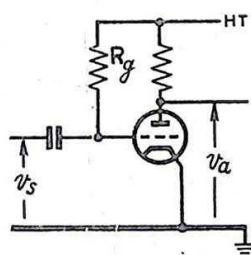


Fig. 329 - Connection providing small positive bias.

Alternatively, the cathode biasing arrangement may be dispensed with, and the grid leak connected to an additional source of negative potential. Although it is not always convenient to employ a separate bias supply voltage, this method has the advantage that the bias is independent of the amplifier valve current.

A few volts positive bias may be obtained by connecting the grid leak to the positive HT supply instead of to earth (Fig. 329). The flow of grid current through R_g should be just sufficient to maintain the grid voltage at the required bias level. For example, suppose the grid-current, grid-voltage conductance is $62 \mu\text{A/volt}$, and the supply voltage is 250 volts. A grid leak of $2 \text{ M}\Omega$ is just sufficient to ensure that the grid is maintained at +2V relative to cathode. The grid current is $124 \mu\text{A}$, which develops across the grid leak 248 volts.

This type of biasing arrangement is suitable for the amplification of negative-going pulses.

In some circuits it is necessary for the mean level of the input voltage to be positive with respect to earth, although the bias developed between grid and cathode may be zero or negative. Examples of this arise in current feedback circuits (Sec. 16) and in the limiter circuits of Chap. 9 Sec. 3.

These are illustrated in Figs. 330 (a) and (b) respectively. The steady voltage at B may be adjusted between the values zero and $\frac{R_2 V_B}{R_1 + R_2}$.

In the arrangement shown in Fig. 331 the steady voltage at B can be adjusted so that it has a value either positive or negative with respect to that of the cathode.

Automatic bias (Slide-Back bias) may be provided by allowing grid current to flow into the coupling condenser in the input circuit of a valve amplifier. The condenser charges when the grid voltage is positive and, in the normal connection illustrated in Fig. 332 (a) discharges when the grid voltage is negative with respect to cathode. In the alternative circuit of Fig. 332 (b), the condenser is discharging continually through the grid leak, the bias adjusting itself so that the

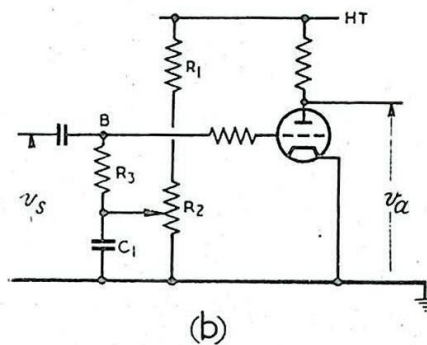
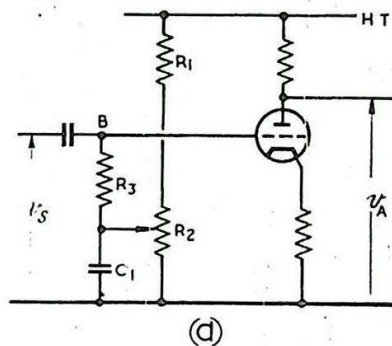


Fig. 330 - Arrangements for providing a positive mean level of input voltage with respect to earth.

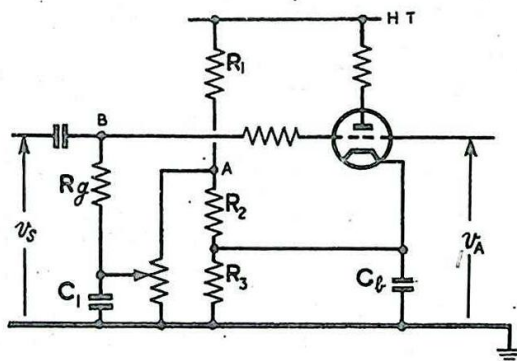


Fig. 331 - Circuit providing positive or negative bias.

more or less constant discharge rate is equal to the mean rate of charging. The principle of this process is dealt with more fully in Clamping Circuits, Chap. 12 Sec. 2. It is also the principle of the Cumulative Grid Detector. Apart from its use as a clamping arrangement or as a detector this method of biasing is most commonly encountered in RF Power Amplifiers and Oscillators. In these circuits distortion introduced by grid current flow is not important, since that due to the use of class C biasing is in any case very considerable. (This distortion is confined to the anode current and not to the output voltage since the anode circuit is normally highly selective and presents a low impedance to all but the first harmonic of the anode current). The principal advantage in such an application is that if the amplitude of the input voltage is reduced the bias is automatically reduced due to a decrease in grid current flow. This causes an increase in the Angle of Flow of anode * current so that the amplitude of the output voltage suffers less reduction than it would if a fixed bias were employed. It is possible for the output amplitude actually to increase as the input amplitude is decreased, due to this selfcompensating action. Where this "Automatic Gain Control" action is detrimental to the action of the amplifier some other method of biasing must be employed.

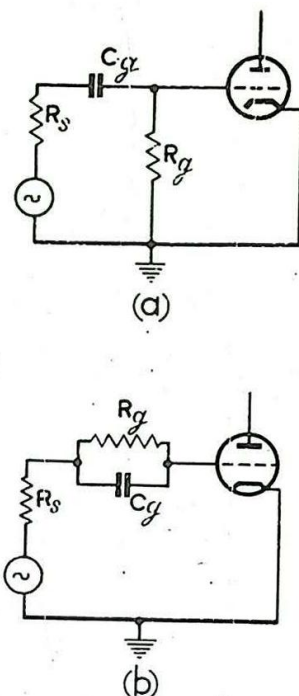


Fig. 332 - Slide-back biasing circuits.

As already pointed out, when slide-back biasing is used with Class C tuned amplifiers and oscillators the additional distortion produced by grid current flow is not normally important. However, there are occasions when the method is applied to resistance-loaded amplifiers and it may be necessary to reduce such distortion to a minimum. This can be arranged by ensuring that the output resistance of the circuit feeding the grid is small, and that the biasing condenser and grid leak are large. If T_1 is the duration of grid current flow, T_2 the period of the applied voltage, R_s the output impedance of the grid input circuit, R_i the input resistance of the valve when grid current flows, R_g the grid leak and C the bias condenser, the following conditions should hold if distortion is to be avoided :-

$$R_i \gg R_s, CR_g \gg T_2, C(R_i + R_s) \gg T_1. \quad (\text{See Fig. 333}).$$

The amount of bias depends on the ratio $\frac{R_g}{R_i + R_s}$. The

larger the ratio the greater is the bias.

* The Angle of Flow is a measure of the duration of current flow in a valve during each cycle of a sinusoidal input voltage. Each period is represented by 360° , and the angle of flow is 360° for Class A, 180° for Class B and correspondingly less for Class C biasing).

If it is desired to introduce distortion without causing appreciable slide-back, either a short time-constant CR_g may be used, producing the type of distortion considered in Chap. 2 Sec. 3, or else the output resistance of the input circuit may be made large whilst the ratio $\frac{R_g}{R_i + R_s}$ is kept

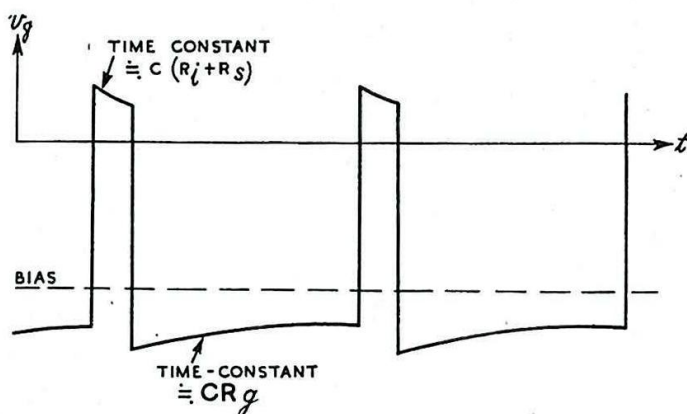


Fig. 333 - Distortion of rectangular pulses by grid current flow.

as low as possible. The limiting action introduced by this arrangement is discussed in Chap. 9 Sec. 3.

5. THE POTENTIAL OF THE SCREEN GRID OF A PENTODE USED AS AN AMPLIFIER

It is generally desired to maintain the screen grid of a pentode amplifier valve at a constant potential. If this potential were not constant but varied with alterations of current through the valve the amplification of the valve would be affected.

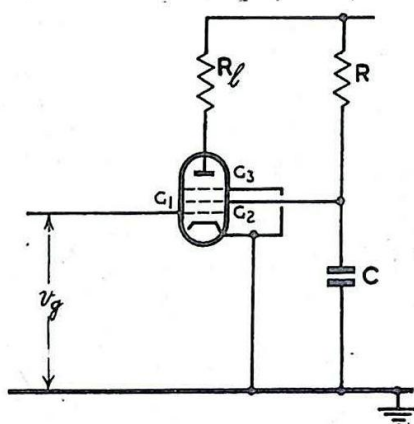


Fig. 334 - Screen supply circuit for class A operation.

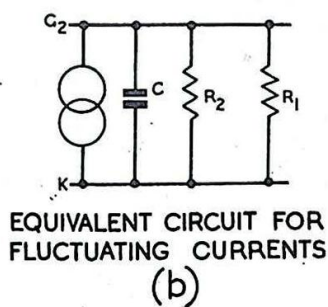
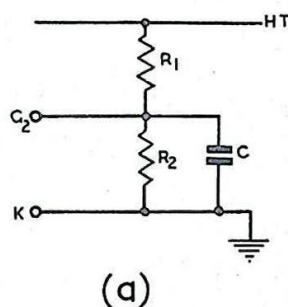


Fig. 335 - Screen supply circuit for class B or C operation.

The simplest method of supplying the screen with a suitable source of voltage is illustrated in Fig. 334. This suffers from the disadvantage of the cathode automatic biasing arrangement, namely that with other than Class A operation the valve current varies with signal amplitude, causing the mean screen potential to rise as the amplitude of the input signal is reduced. For such conditions the arrangement of Fig. 335 is to be preferred.

The condenser in Figs. 334 and 335 is adequate for smoothing out the fluctuation in screen potential provided the time-constant of the decoupling circuit is large compared with the period of the input voltage. In Fig. 334 this time-constant is CR . In Fig. 335 the resistors R_1 and R_2 are effectively in parallel so far as fluctuating currents are concerned as illustrated at (b), so that the effective time-constant is $\frac{C R_1 R_2}{R_1 + R_2}$.

6. ANODE DECOUPLING OF AMPLIFIERS

It has so far been assumed that the HT supply circuit presents a negligible impedance to alternating currents. This is equivalent to the assumption that the time-constant of the supply is very long, and is justified if the smoothing condenser in the power pack is sufficiently large. In practice the voltage across the HT supply circuit varies to a certain extent in response to fluctuations of current passing through this circuit. These fluctuations are present in the current passing through a valve and so pass to the HT supply. A fluctuating voltage is, therefore, developed across the HT supply circuit and so is applied to other amplifier stages using this supply. This feedback of the fluctuating voltages is undesirable, as it may, particularly if applied to a high gain amplifier, cause continuous oscillations to be set up.

The method which is usually adopted to avoid the feedback described above is to prevent the variations of anode currents of the amplifier valves from flowing through the HT supply circuit, by the use of an anode decoupling circuit. Fig. 336 shows the anode decoupling circuit for one amplifier, the circuit consisting of a resistor R and condenser C . Provided the time-constant CR is long compared with the period of the applied voltage, the potential of the point A does not change appreciably with variations of the applied voltage. The point A is maintained at steady potential $R\bar{i}_a$ below HT level where \bar{i}_a is the steady anode current of the valve about which the fluctuations of current occur. The introduction of the decoupling circuit necessitates an increase of HT potential by an amount $R\bar{i}_a$ if the mean anode potential is to remain the same as it was without the decoupling circuit.

Where a very considerable degree of decoupling is required it is usual to employ two or more C-R networks in cascade as shown in Fig. 337. By this arrangement a high degree of smoothing may be achieved without the use of prohibitively large resistors or condensers. For example, suppose that it is necessary to prevent more than 1% of the valve current fluctuations from reaching the supply circuits. This may be provided by a single circuit with time-constant T , say. The same effect may be obtained by the use of two circuits in cascade, each having a time-constant $\frac{T}{10}$, so that smaller resistors and condensers may be employed.

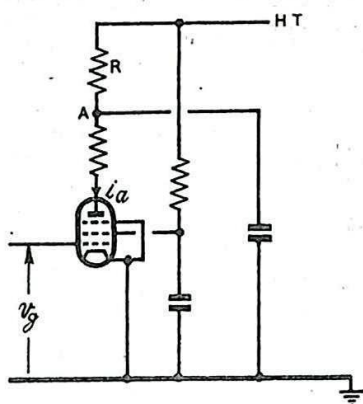


Fig. 336 - Anode de-coupling circuit.

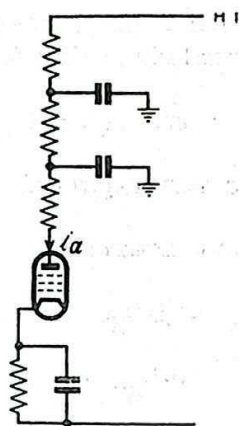


Fig. 337 - Decoupling networks in cascade.

7. MILLER EFFECT

If a triode is operating in an amplifier circuit the interelectrode capacitance C_{ga} of the valve between grid and anode provides coupling between the output and input circuits. This coupling affects the input admittance of the valve, the precise effect depending upon the type of load present in the output circuit.

Fig. 338 (a) shows a typical amplifier circuit, the interelectrode capacitances between grid and anode and between grid and cathode being denoted by C_{ga} and C_{gk} respectively.

The voltage across C_{ga} , as shown in the equivalent circuit of Fig. 338 (b) has an alternating component

$$\begin{aligned} v_i - v_o &= v_i - \mu v_i \\ &= v_i (1 - \mu). \end{aligned}$$

Hence the current i_1 through C_{ga} is given by

$$i_1 = j\omega C_{ga} \cdot v_i (1 - \mu).$$

The current i_2 through C_{gk} is given by

$$i_2 = j\omega C_{gk} v_i.$$

Hence the total input current i_1 is given by

$$i_1 = j\omega v_i \left[C_{gk} + (1 - \mu) C_{ga} \right].$$

The input admittance is therefore

$$Y_i = \frac{i_1}{v_i} = j\omega C_{gk} + j\omega C_{ga} (1 - \mu).$$

If θ is the phase displacement between the grid and anode voltages, the voltage amplification m can be expressed in the form

$$m = |m| (\cos \theta + j \sin \theta)$$

where $|m|$ is the magnitude of the amplification.

Hence the input admittance can be expressed as

$$\begin{aligned} y_i &= j\omega C_{gk} + j\omega C_{ga} (1 - |m| \cos \theta - j |m| \sin \theta) \\ &= \omega C_{ga} |m| \sin \theta + j (\omega C_{gk} + \omega C_{ga} (1 - |m| \cos \theta)) \end{aligned}$$

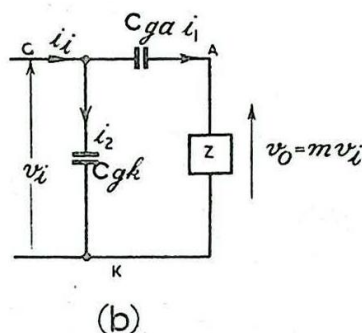
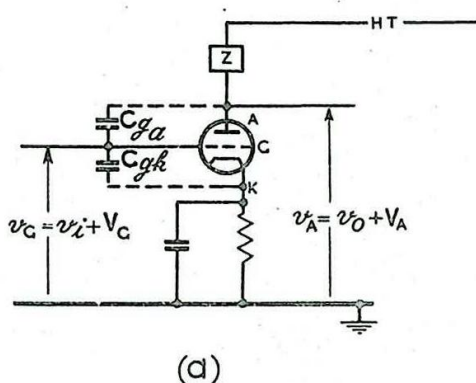


Fig. 338 - Miller Effect.

Provided the effective load (i.e. load impedance in parallel with stray and interelectrode capacitance) of the amplifier is purely resistive, the grid and anode voltages are approximately anti-phase and hence $\theta \approx 180^\circ$. Under these circumstances the input admittance is a capacitive susceptance of value

$$\omega C_{gk} + \omega C_{ga} (1 + |m|).$$

Consequently the effective input capacitance of the valve is equal to

$$C_{gk} + C_{ga} (1 + |m|).$$

For a pentode C_{ga} is very small (less than 0.5 per cent of C_{gk}). Consequently the effective input capacitance of a pentode employed in an amplifying circuit is likely to be considerably smaller than that of a triode ($C_{ga} \approx C_{gk}$) in spite of large values of m corresponding to the pentode.

If the load of an amplifier is inductive θ is greater than 180° and less than 270° . Hence $\sin \theta$ and $\cos \theta$ are negative and the input admittance is due to a negative conductance ($\omega C_{ga} |m| \sin \theta$) and a capacitance susceptance ($\omega C_{gk} + \omega C_{ga} (1 - |m| \cos \theta)$).

In the case of a valve employed in an RF amplifier circuit the anode load usually consists of a tuned circuit. For some frequencies off resonance this load is inductive; consequently the input conductance is negative and regeneration is likely to occur, possibly leading to continuous oscillations. If a triode is

used, C_{ga} , and hence the value of the negative conductance, is large. Normally if a triode is used in an RF amplifier circuit neutralisation of the effect due to feedback via C_{ga} is necessary. Neutralisation may be performed if a voltage in the output circuit, equal in amplitude and anti-phase to the anode voltage, is impressed on the grid of the valve through a capacitance approximately equal to C_{ga} . Fig. 339 shows a circuit employing neutralisation.

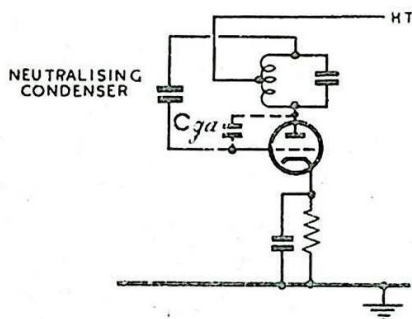


Fig. 339 - Neutralising circuit used in Power Amplifiers.

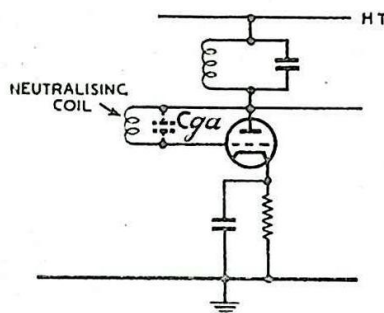


Fig. 340 - Neutralising circuit used in Radio Receivers.

Alternatively, if a coil of suitable inductance is connected between anode and grid so that it forms with C_{ga} a parallel circuit resonant at the frequency of operation, negligible feedback occurs; (Fig. 340).

If the load of the amplifier is capacitive, θ is greater than 90° and less than 180° . Hence $\sin \theta$ is positive. Under these circumstances the input conductance ($\omega C_{ga} \sin \theta$) is positive, and damping is introduced into the input circuit of the amplifier.

8. RF AMPLIFICATION

The applications of RF amplifiers to the problems of radio reception are dealt with in Chap. 16. Here we are concerned only with the types of amplifying circuits which may be used.

These circuits are essentially the same whether used in the IF or RF stages of a receiver, the chief factor in design being the ratio of the bandwidth to the mid-frequency. The ideal amplification-frequency and phase-frequency characteristics for such an amplifier are illustrated in Fig. 341. Throughout the required band of frequencies, the amplification should be uniform and any phase-shift introduced should be proportional to $f - f_m$ where f is the frequency and f_m the mid-frequency of the band.

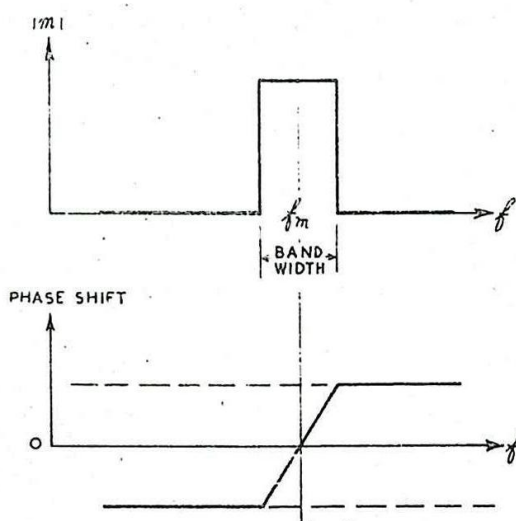


Fig. 341 - Idealised Characteristics of a R.F. Amplifier.

For amplifying RF pulses of about one micro-second duration the bandwidth is usually of the order of 2-4 Mc/s, whilst the mid-frequency may have any value from 9 Mc/s to 600 Mc/s. Amplification at frequencies higher than this is not normally attempted.

The most usual type of circuit is the simple tuned amplifier arrangement, illustrated in Fig. 342 (a). The equivalent circuit is shown at (b).

r_{oD} is the dynamic resistance of the tuned circuit at resonance (Fig. 342 (c)).

lR_D is the equivalent dynamic resistance of the amplifier load circuit, consisting of R_a , r_{oD} and R_l in parallel.

It follows that at resonance the amplification $\frac{\hat{v}_o}{\hat{v}_i}$ is equal to $G_m lR_D$.

The amplification at frequencies other than the resonant frequency is proportional to the impedance of the resultant circuit (Fig. 342 (b)) and follows the normal response curve for a parallel tuned circuit. The selectivity, or Q for this curve is Q_l , and is related to Q_o of the undamped tuned circuit by the equation

$$\frac{Q_l}{lR_D} = \frac{Q_o}{r_{oD}}$$

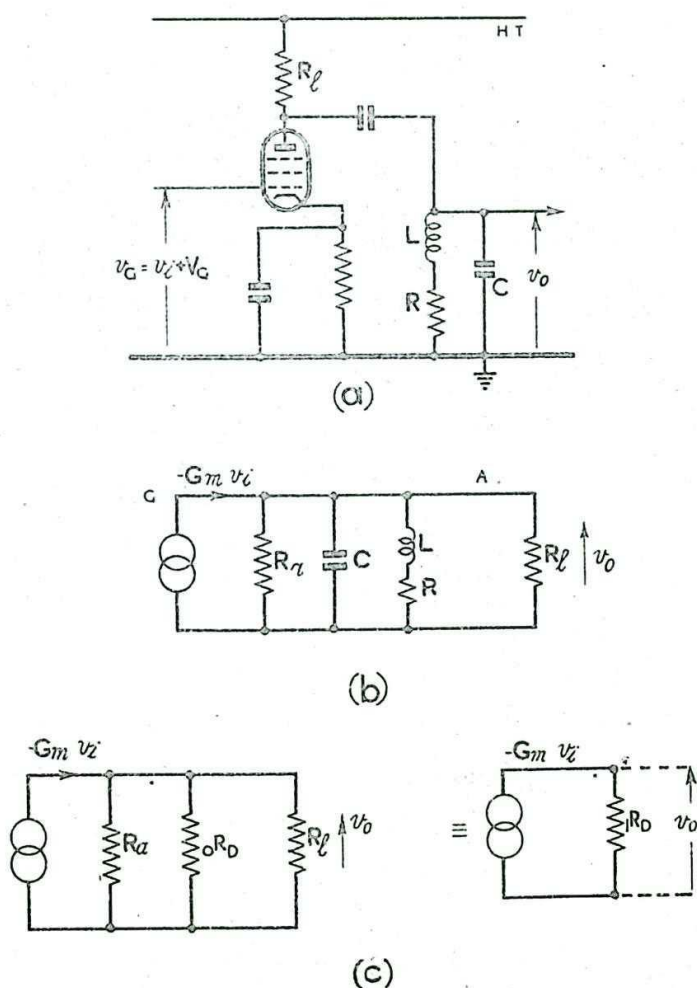


Fig. 342 - R.F. amplifier circuit.

Above about 30 Mc/s the input capacitance of the succeeding stage together with the self-capacitance of the RF coil provides sufficient capacitance for resonance. At such frequencies Permeability Tuning may be employed. This consists of inserting a moveable core either of copper or of powdered iron. The eddy currents set up in the copper reduce the effective inductance of the tuned circuit. The high permeability of the iron core, powdered to prevent losses due to eddy currents, increases the effective inductance.

At frequencies of the order of 200 Mc/s and upwards lumped L-C circuits are often replaced by tuned lines, usually coaxial with plunger tuning. At frequencies above about 30 Mc/s the input resistance of a valve amplifier is small, so that the equivalent circuit of Fig. 342 (b) is shunted by the low input resistance of the succeeding stage. Normally R_a and R_p (Fig. 342 (c)) are of the order of $1\text{ M}\Omega$ and $100\text{ K}\Omega$ respectively; R_f is usually only a few thousand ohms, this fact by itself accounting for wide bandwidth and low gain. The low value of the input resistance further heightens this effect. This resistance decreases with frequency (see Sec. 25) so that the higher the frequency the wider is the bandwidth and the lower the gain.

At lower frequencies where the input resistance is not important, higher gains can be achieved by the use of an RF choke in place of R_f .

An alternative circuit for an RF amplifier is the series-fed arrangement shown in Fig. 343. In the absence of a low input resistance, higher amplification can be obtained from this circuit than that of Fig. 342 (a) since R_g is normally much greater than R_f .

Because of the heavy damping of the tuned circuit at high frequencies it is essential to use a valve with a large mutual conductance if appreciable amplification is to be provided. Practical values for the G_m of typical RF pentodes are 6.5 mA/V (6V1091) and 8.5 mA/V (6V1065).

Substantially uniform amplification and linear phase-shift over a wide band of frequencies is often obtained by band-pass coupling: Fig. 344; (See also Chap. 1 Sec. 21). If the resonant frequencies of the two tuned circuits are equal the pass band is determined by the coefficient of coupling.

In some cases double-tuned transformers are employed, the circuit appearing as in Fig. 344, but primary and secondary circuits being tuned to different frequencies on either

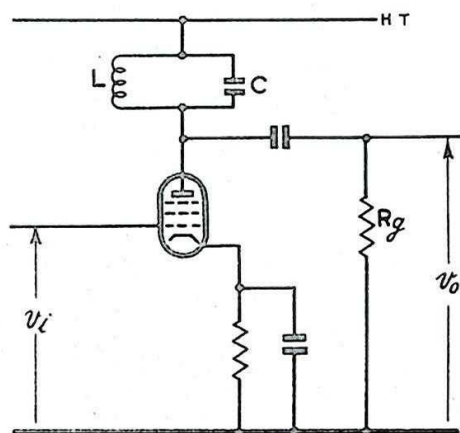


Fig. 343 - Alternative (series-fed) R.F. amplifier circuit.

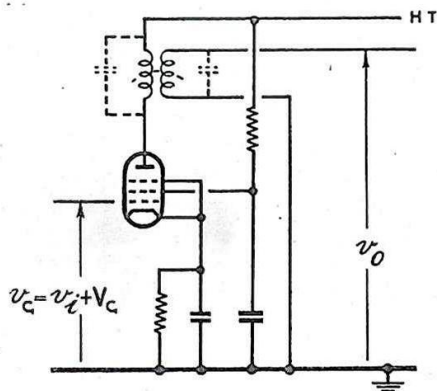


Fig. 344 - Band-pass Circuit.

side of the mid-frequency of the amplification band. The coupling in this case is not so tight as that for the former circuit, and some loss of amplification occurs. The advantage lies in ease of adjustment, since with less tight coupling the tuning of primary and secondary circuits is less interdependent.

A common type of circuit arrangement consists of several pairs of stages using single tuned circuits, each stage of a pair being tuned to a frequency differing from the mid-frequency by an amount equal but in the opposite direction to the other. Such an arrangement is known as Staggered Tuning. Thus with five stages where the intermediate frequency is 7 Mc/s, the first and third stages might be tuned to 5.5 Mc/s, the second and fourth to 8.6 Mc/s and the fifth to 7 Mc/s. The overall response curve then indicates uniform amplification and linear phase-shift over a frequency band of about 3 Mc/s. (Fig. 345).

In general, for a given number of stages employing single tuned circuits it is much more economical to obtain the necessary bandwidth by means of staggered tuning than by resistive damping alone, the latter arrangement resulting in very low gain, and the response curve being far from ideal. For ease of lining-up the amplifier, i.e., adjustment of the various tuned circuits, it is preferable that all the stages should be tuned to the same frequency. Staggered tuning is next in order of simplicity, and the use of double-tuned transformers or tightly coupled circuits is still more complicated, due to interaction between primary and secondary circuits.

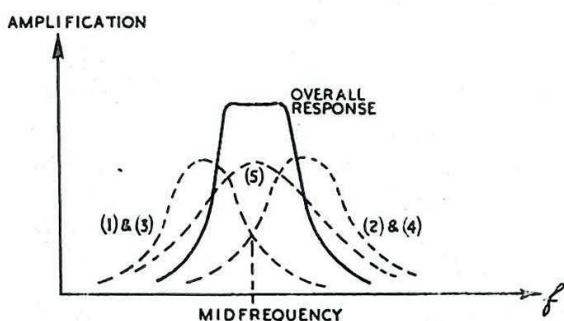


Fig. 345 - Response curve: staggered tuning.

VIDEO-FREQUENCY AMPLIFICATION

9. General Considerations

In a normal broadcasting receiver it is common to refer to post-detector amplification as Audio-Frequency amplification, since the signal to be amplified can be analysed into Fourier components with frequencies in the audible range from 50 c/s to about 4000 c/s. The detected signal of a radar receiver, on the other hand, normally consists of a train of rectangular pulses of short duration (1 to 3 μ s) and with a repetition frequency of the order of 400 to 4000 c/s.

Fourier analysis shows that a succession of rectangular pulses can be considered as consisting of sinusoidal components covering a range of frequencies which in theory is infinite. In practice, frequencies up to a few megacycles only are important, and higher components may be ignored. Signal voltages resolvable into components within this range of frequencies are also common in television practice, and since they constitute the intelligence which goes to make the picture they are usually referred to as Video-Frequency signals in contrast to the Audio-Frequency signals of sound-reproducing systems.

This nomenclature has been carried over to radar, and the amplification of voltage pulses is referred to as Video-Frequency Amplification.

A video-frequency amplifier should provide uniform amplification without phase-shift for voltages of frequencies up to several megacycles. It is difficult in practice to avoid frequency and phase distortion in such an amplifier, and, as discussed below, it is necessary to be content with low gain in order to obtain a satisfactory response.

It is usual for a video-frequency amplifier to have a resistive load, since this has, theoretically, uniform response to all frequencies. In practice stray capacitance in parallel with the resistance may interfere with the performance.

Amplitude distortion is always present in amplifiers, although it can be reduced in magnitude by careful circuit design. Some video-frequency amplifiers are operated under such conditions that non-linear portions of the valve characteristics are used deliberately. (See Chap. 9).

10. The Amplification of an Instantaneous Change of Voltage

If an instantaneous change of voltage (Fig. 346 (a)) is applied to the grid of a valve, which is biased so that it operates on the linear part of its dynamic characteristic, (assuming the load to be purely resistive, as shown in Fig. 347), then the anode current in the valve varies in a manner similar to that of the applied voltage (Fig. 346 (b)).

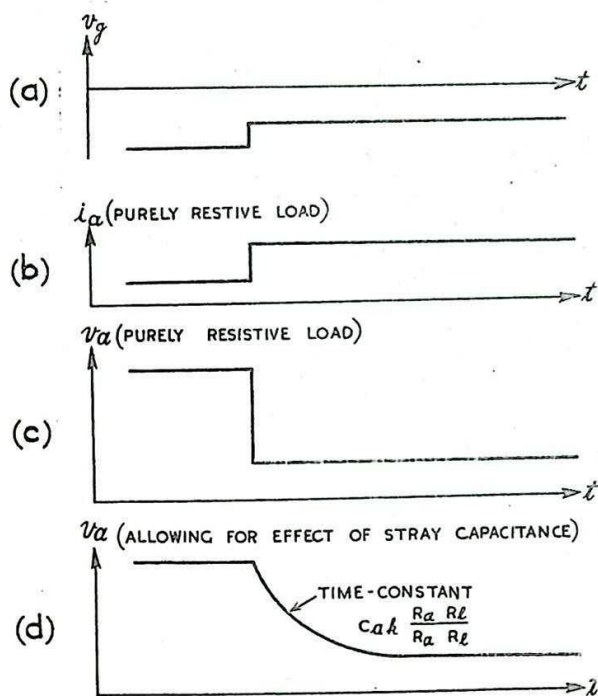


Fig. 346 - Effect of stray capacitance on output voltage.

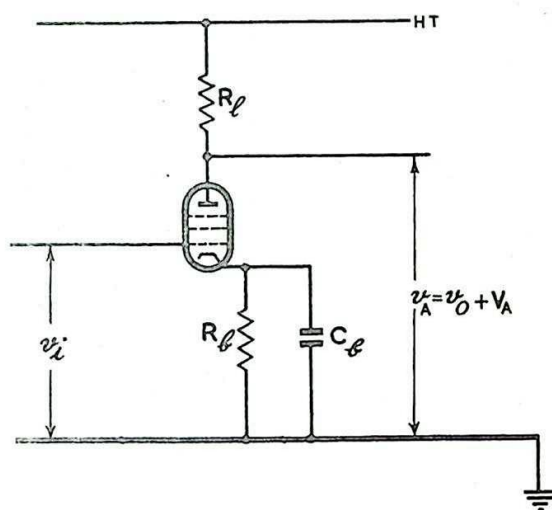


Fig. 347 - Video-frequency amplifier.

The variation of anode potential is exactly inverse to that of the anode current (Fig. 346 (c)). In fact, however, there are usually stray capacitances between anode and earth, so that the anode load consists

effectively of a resistor and condenser in parallel (Fig. 348).

The equivalent circuit of the amplifier shown in Fig. 348(a) is given at (b). To consider the response of this circuit to an instantaneous change v_i of the input voltage between grid and cathode we rearrange this in the constant-voltage form as shown in Fig. 348(c) (See Chap. 1 Sec. 21). The response of this circuit to an instantaneous change of voltage has already been dealt with in Chap. 2. In this case the effective generator voltage is

$$-G_m R_l v_i, \text{ where } R_l$$

$$= \frac{R_a R_l}{R_a + R_l}, \text{ and the}$$

response is shown in Fig. 346 (c). The time-constant of the exponential change of voltage at the anode is $C_{ak} R_l$.

The change of anode potential corresponds more and more closely to the change of grid potential as the time-constant $C_{ak} R_l$ is diminished, by reducing either R_l or the capacitance C_{ak} . Reduction of R_l , however, results in a decrease in amplification.

If the output from valve 1 (Fig. 349 (a)) is applied to valve 2 through a capacitance C and resistance R_g , with C_{gk} representing the input capacitance of valve 2, then the equivalent circuit for valve 1 has the form shown in Fig. 349 (b). The voltage v_o produced across the circuit between 2G and earth is applied to the grid of valve 2. Since the current through the circuit changes very rapidly (ideally instantaneously) the condenser C , which is assumed to have a capacitance large compared with the stray capacitances, can neither charge nor discharge appreciably. Therefore, so far as rapid changes in the circuit are concerned, the effective equivalent circuit is as shown in Fig. 349 (c). The voltage v_o falls exponentially by an amount

$$\frac{R_l R_g}{R_l + R_g} \cdot G_m \hat{v}_i,$$

with a time-constant

$$\frac{R_l R_g}{R_l + R_g} (C_{ak} + C_{gk}).$$

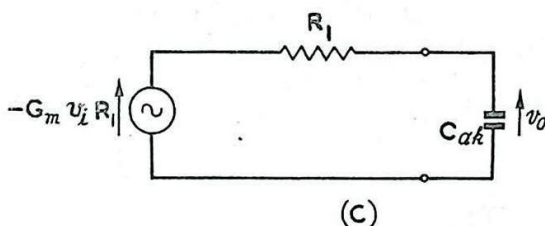
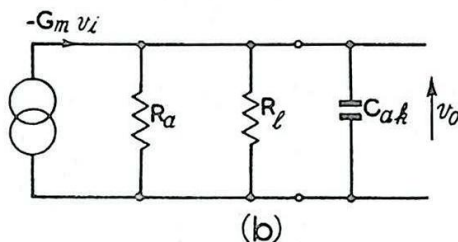
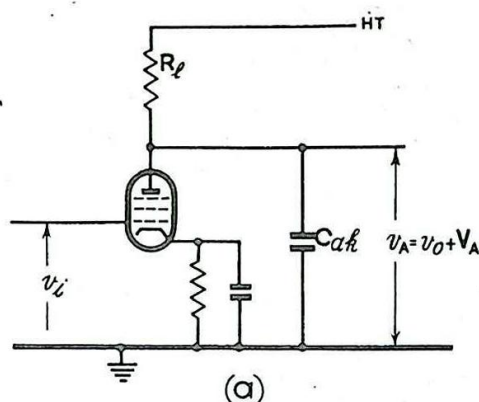


Fig. 348 - Effect of stray capacitance on video-frequency amplification.

In practice, the capacitance C_{gk} is usually much greater than C_{ak} and is therefore the controlling factor in this time-constant.

C_{gk} is assumed to include any input capacitance of valve 2 arising from the Miller effect (Sec. 7). The condenser C now discharges exponentially through R_g and

R_1 in series (Fig. 349 (b)) with time-constant $C(R_g + R_1)$ until the total fall in anode voltage becomes $R_1 G_m \hat{v}_i$.

Meanwhile the grid voltage rises towards zero. The total changes of voltage at the anode of valve 1 and the grid of valve 2 are illustrated in Fig. 350.

If C is not very large compared with C_{gk} these two capacitances have a potential dividing effect, and the amplitude of the initial fall in voltage at the grid of the second valve is reduced.

(See Chap. 2 Sec.5). This is likely to occur only when a short time-constant coupling circuit is used for the purpose of distorting the pulse-shape.

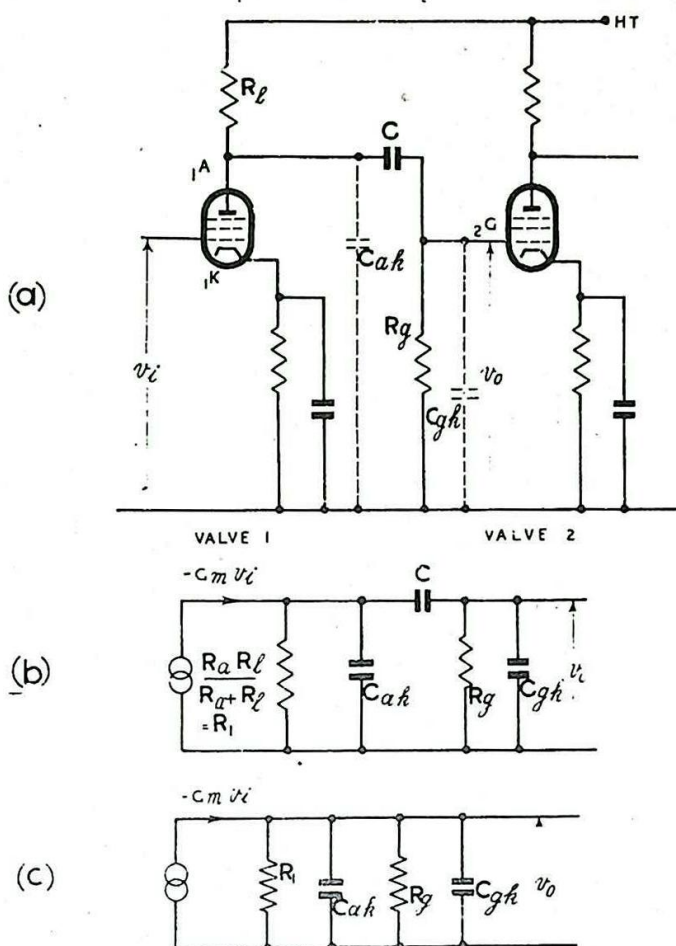


Fig. 349 - Two-stage video-frequency amplifier.

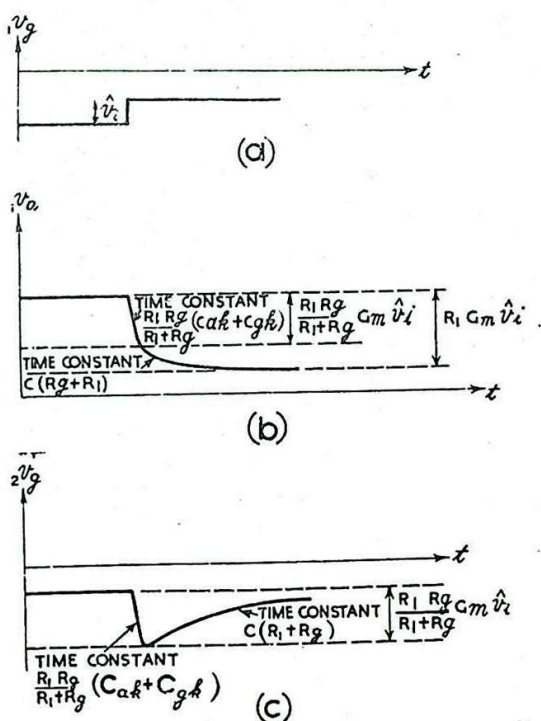


Fig. 350 - Effect of stray and coupling capacitances on video-frequency amplification.

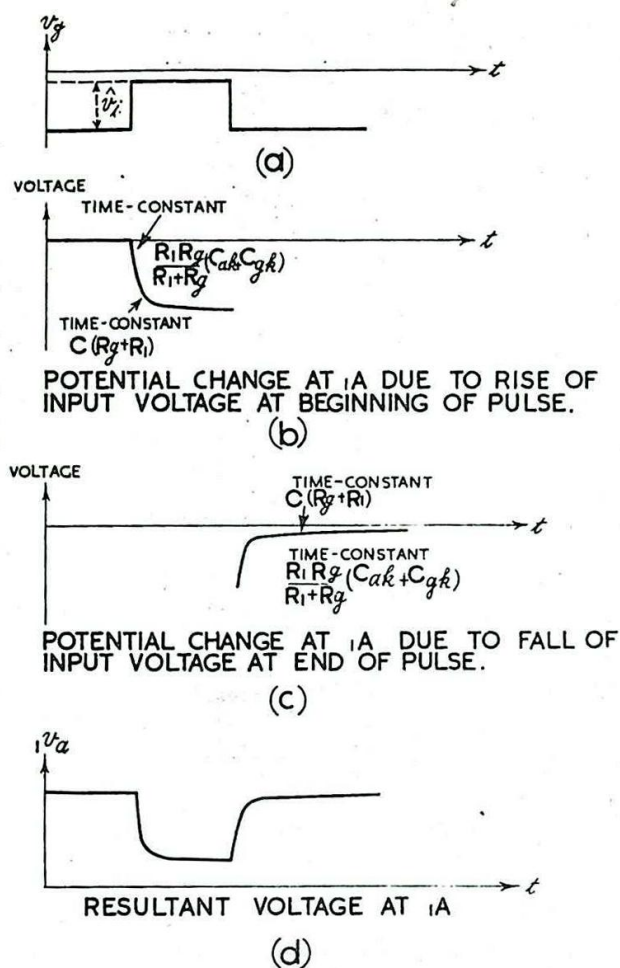


Fig. 351 - Response of a video-frequency amplifier to a rectangular pulse at the grid.

11. The Amplification of a Rectangular Voltage Pulse

The response of a video-frequency amplifier to a rectangular voltage pulse applied between grid and cathode may be found by considering the effect of applying a sudden change of voltage followed after an interval by another sudden change of equal amplitude but of opposite polarity.

If the pulse shown in Fig. 351 (a) is applied to the circuit shown in Fig. 349 the resultant changes in potential at the anode of valve 1 and the grid of valve 2 are as shown in Figs. 351 and 352 respectively. The waveform of Fig. 351 (d) is obtained by adding to the initial potential of the anode of valve 1 the changes of potential at the anode due in turn to the negative-going and positive-going changes illustrated in Fig. 351 (b) and (c). The waveform of Fig. 352 (c) is obtained by adding to the initial potential of the grid of valve 2, the potential changes of the grid due in turn to the negative-going and positive-going changes illustrated in Fig. 352 (a) and (b).

The time-constant $\frac{R_g R_1}{R_1 + R_g} \cdot (C_{ak} + C_{gk})$ determines the rates of

rise and fall of the output pulse, whilst the time-constant $C \cdot (R_1 + R_g)$

determines the constancy of the output voltage during the period of the pulse. The potential produced at the grid of valve 2 approximates most closely in form to the potential applied to the grid of valve 1 if the time-constant $C(R_1 + R_g)$ is very long compared with the duration

of the input pulse and if the time-constant $\frac{R_g \cdot R_1}{R_g + R_1} (C_{gk} + C_{ak})$ is

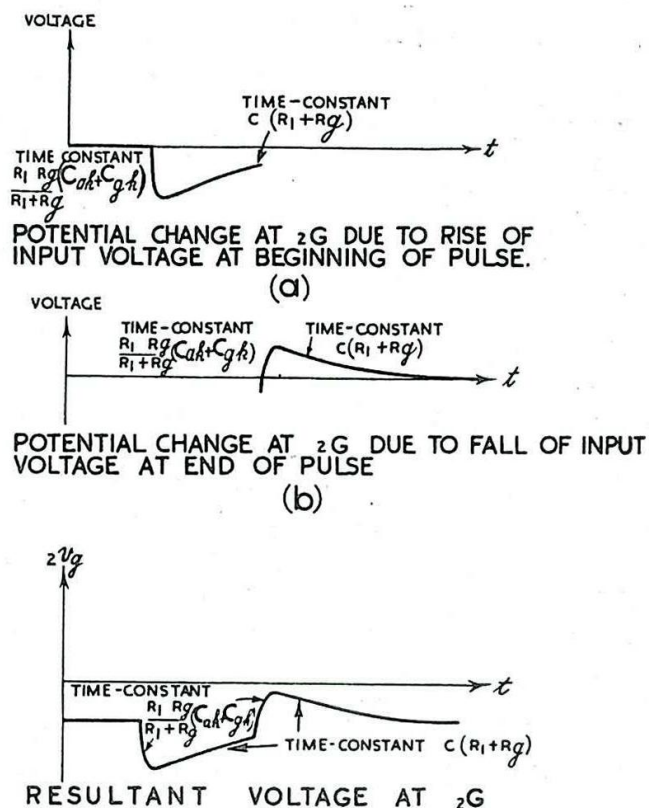


Fig. 352 - Response produced at the grid of valve 2 by rectangular pulse at grid of valve 1.

very short compared with the duration of the input pulse. It is assumed, as is normally the case, that $C \gg C_{gk}$.

The pulse developed between the grid and cathode of valve 2 is in the opposite sense to that applied between the grid and cathode of valve 1. This corresponds to the change of phase of 180° obtained when a sinusoidal voltage is amplified by a valve with a purely resistive load.

The amplification of the applied voltage pulse by valve 1 is

$$= \frac{G_m \cdot R_1 R_g}{R_1 + R_g}$$

Generally, R_g is made large compared with R so that the total amplification is approximately $G_m \cdot R_1$, or, with a pentode, $G_m R_1$, since $R_g \gg R_1$.

The amplification of the circuit as a whole is increased by increasing R_1 , but if this is done the time-constant $\frac{R_g \cdot R_1}{R_g + R_1} \cdot (C_{ak} + C_{gk})$ is increased so that the rise and fall of the output pulse

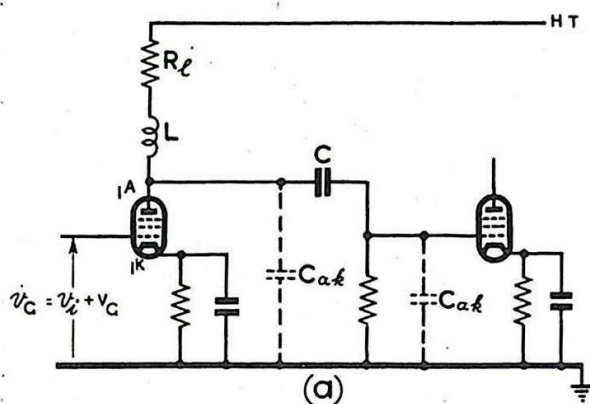
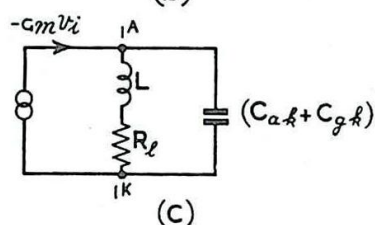
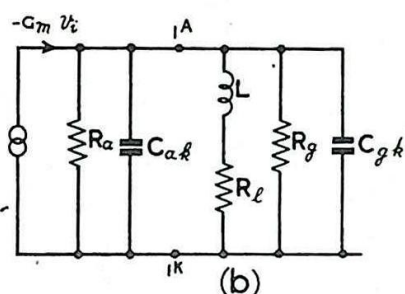


Fig. 353 - Use of compensating coil in anode circuit.



are less rapid. For example suppose $R_g \gg R_L$, and $C_{gk} + C_{ak} = 20$ pF. If the rise or fall of the output pulse is to be completed in $\frac{1}{10}$ μ s, then $5 \times (\text{time-constant of rise or fall of output voltage}) < \frac{1}{10}$ μ s; i.e.,

$$5 \cdot \frac{R_g \cdot R_L}{R_g + R_L} (C_{ak} + C_{gk}) < \frac{1}{10} \mu\text{s}$$

$$\text{or } 5 \cdot R_L \cdot 20 \cdot 10^{-12} < \frac{1}{10} \cdot 10^{-6}$$

$$\text{therefore } R_L < 1000 \text{ ohms}$$

or in the case of a pentode $R_L < 1000$ ohms.

For a pentode of mutual conductance 4 mA/volt the amplification $G_m R_L$ cannot be greater than $0.004 \times 1000 = 4$.

For a given value of R_L , the rise and fall of the output pulse can be made more rapid by introducing a coil of inductance L in series with the anode load resistor as shown in Fig. 353 (a). The equivalent circuit for rapid changes of voltage is then as shown in Fig. 353 (b). If R_a and R_g are large compared with R_L , the circuit of Fig. 353 (b) approximates to that of Fig. 353 (c). The circuit may act as an over-damped "ringing circuit". If L is suitably chosen the response of the anode circuit when shock excited by a sudden change

of valve current is more rapid than that of a circuit consisting of R_f and $(C_{ak} + C_{gk})$ alone.

(See Chap. 1, Sec. 21).

If the condition

$$L = \frac{1}{2} R_f^2 (C_{ak} + C_{gk})$$

is satisfied, it can be shown that the time taken for the output pulse to rise or fall is $\frac{3\pi}{8} R_f (C_{ak} + C_{gk})$,

compared with $5 R_f (C_{ak} + C_{gk})$ for the circuit without the coil. Fig. 354 shows the fall of anode voltage, compared with that of the uncompensated circuit.

If the rise or fall of the output pulse is to be complete in $\frac{1}{10} \mu s$,

then

$$\frac{3\pi}{8} \cdot R_f \cdot 20 \text{ pF} = \frac{1}{10} \mu s$$

Hence, $R_f = 5000 \text{ ohms approximately,}$

and
$$L = \frac{1}{2} R_f^2 (C_{ac} + C_{gc}) = 250 \mu H.$$

For a pentode of mutual conductance 4 mA/volt the amplification $G_m R_f$ is equal to $0.004 \cdot 5000 = 20$.

The addition, therefore, of a coil of small inductance in series with the anode load of a pulse amplifier allows the amplification of the circuit to be increased without loss of steepness of the rise and fall of the output pulse.

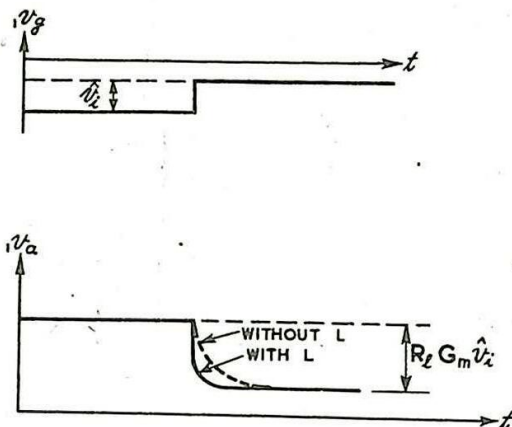


Fig. 354 - Effect of compensating coil in anode circuit.

12. The Frequency Spectrum Approach to Video-Frequency Amplification

The response of a video-frequency amplifier to voltage pulses may be calculated by means of methods discussed in Chap. 16 Sec. 1. The Frequency Spectrum of the input voltage is found by harmonic analysis, and then the effect of the amplifier circuit on the phase and amplitude of the components at the various frequencies is calculated. The mathematical technique involved is complicated. It is, however, comparatively simple to discover the conditions which produce distortion even if the determination of the exact form of the output voltage is difficult.

The variation of amplification and phase-shift with Frequency for the circuit of Fig. 349 (a), shown in its equivalent form at (b), may in general be formed by dividing the frequency range into three parts.

- (i) Over the mid-frequency region the shunt susceptance $\omega(C_{ak} + C_{gk})$ is sufficiently small and the series susceptance ωC sufficiently large to be negligible in comparison with the conductances. Hence the equivalent circuit takes the form shown in Fig. 355 (a).
- (ii) At high frequencies the series susceptance is still negligible but the shunt susceptance is comparable with the conductance; (b).
- (iii) At low frequencies the shunt susceptance is negligibly small, but the series susceptance is comparable with the conductance; (c).

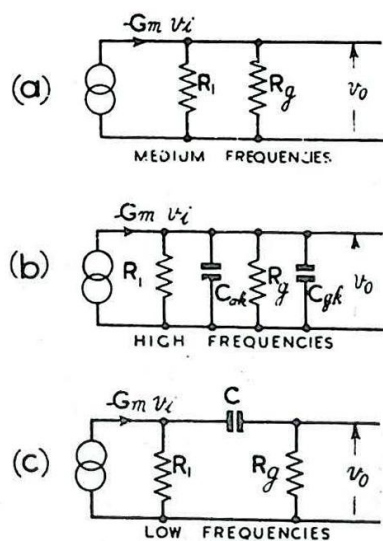


Fig. 355 - Equivalent circuits for video-frequency amplifier.

At medium frequencies the amplification is

$$G_m \frac{R_1 R_g}{R_1 + R_g}$$

and the phase-shift by which the input leads the output is 180° .

At high frequencies the amplification is approximately $G_m \frac{R_1 R_g}{R_1 + R_g}$

$\cos \phi_h$ and the phase-shift is $180^\circ + \phi_h$, where ϕ_h is given by

$$\tan \phi_h = \omega \frac{R_1 R_g}{R_1 + R_g} (C_{ak} + C_{gk}).$$

For example, the input leads the output by 225° where $\omega =$

$$\frac{R_1 R_g}{R_1 R_g (C_{ak} + C_{gk})} \quad \text{so that } \tan \phi_h = 1.$$

The amplification is $\frac{1}{\sqrt{2}}$ times that at medium frequencies, representing a reduction in voltage gain by 3 db.

At low frequencies the amplification is approximately $G_m \frac{R_1 R_g}{R_1 + R_g} \cdot \cos \phi_l$ and the phase-shift is $180^\circ - \phi_l$, where ϕ_l is given by

$$\tan \phi_l = \frac{1}{\omega C (R_1 + R_g)}$$

For example, the input leads the output by 135° when $\omega = \frac{1}{C (R_1 + R_g)}$

so that $\tan \phi_l = 1$. As before, the amplification is $\frac{1}{\sqrt{2}}$ times that

at medium frequencies. The overall amplification-frequency and phase-frequency characteristics for a typical video-frequency amplifier are shown in Fig. 356, plotted to a logarithmic frequency scale.

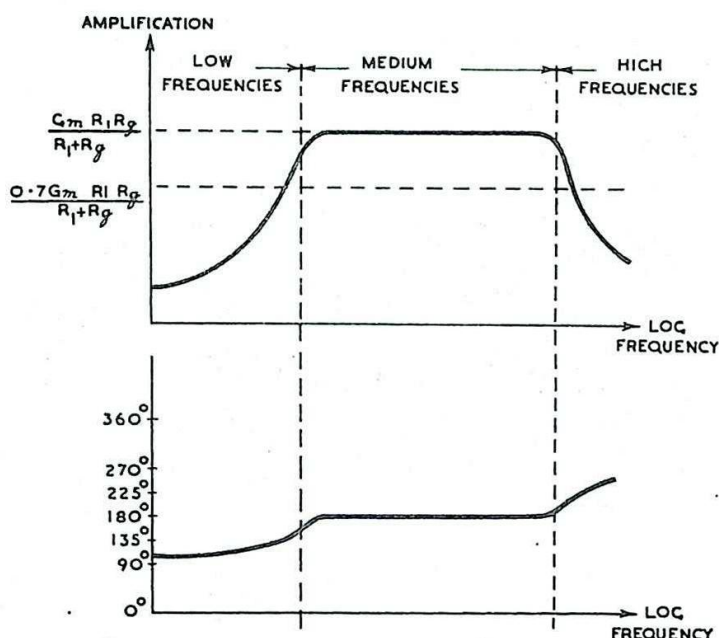


Fig. 356 - Amplification and phase shift characteristics for a video-frequency amplifier.

13. The Use of Pentodes in Video-Frequency Amplification.

Pentodes are normally preferred to triodes in video-frequency amplifiers because of the greater amplification obtainable.

This is due to the fact that R_1 , which equals $\frac{R_a R_l}{R_a + R_l}$, can be

made much larger for a pentode with an R_a of the order of megohms, than for a triode, for which R_a is a few thousand ohms. The time-constant for rapid changes in anode voltage is thereby made larger by the increase in R_1 . However, a pentode normally has a smaller input capacitance than a triode, so that when two stages of amplification are considered the increase in the time-constant due to the larger value of R_1 for the first stage is offset by the reduction in input capacitance of the second stage. Thus the use of pentodes enables greater

amplification to be obtained without an increase of distortion, compared with that due to triodes.

NEGATIVE FEEDBACK AMPLIFICATION

14. General Principles

A feedback amplifier is one in which a fraction, or all, of the output voltage or current of an amplifier is fed back via a coupling network to the input circuit of the amplifier.

Consider the case of an undistorting amplifier, with a feedback network which also produces no distortion. The feedback is termed positive if the magnitude of the input voltage is increased, and negative if it is decreased by the feedback connection. In both cases the properties of the amplifier are modified as regards amplification, input and output impedances and distortion.

In general, some form of distortion is invariably produced either by the amplifier or by the feedback network, and it is convenient to consider the nature of the feedback for sinusoidal input voltages at various frequencies. Usually over some range of frequencies the fed-back voltage is largely in quadrature with the input, whilst it is possible for positive feedback (regeneration) to occur in a circuit in which negative feedback predominates over the principal frequency range.

Suppose a sinusoidal voltage v_s is applied to the input terminals (1) and (2) of the amplifier shown in Fig. 357. Let v_i be the voltage developed between terminals (3) and (4). If m is the amplification of the amplifier alone, the output voltage v_o is equal to $m v_i$.

Since a fraction β of v_o is fed back to the input, the voltage v_i is given by $v_s + \beta v_o$ and the output v_o produced by this input is $m(v_s + \beta v_o)$.

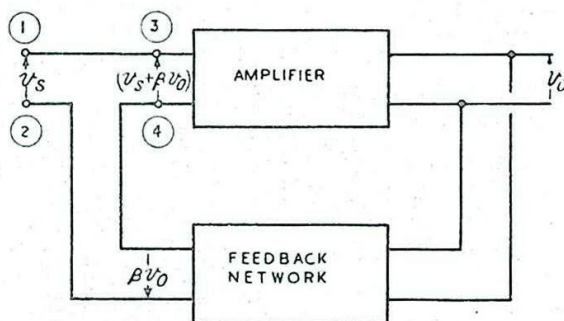


Fig. 357 - Schematic diagram showing feedback.

$$\text{Therefore } m(v_s + \beta v_o) = v_o$$

$$\text{so } v_o = \frac{m v_s}{1 - \beta m} \dots \dots \dots (1)$$

In general m and β are both complex, amplification with and without feedback being accompanied by phase-shift which varies with frequency. In the particular case where β and m are real, as, for example where the feedback network is resistive and the amplifier is a video-frequency amplifier operating over its mid-frequency range, we may deduce simple expressions for the overall amplification and the effect on amplitude distortion.

In such a case, equation 1 gives

$$\frac{v_o}{v_s} = \frac{m}{1 - \beta m} \quad (\text{where } m \text{ is negative}) \dots\dots\dots(2)$$

For negative feedback β is positive and the amplification is reduced in the ratio $\frac{1}{1 - \beta m}$. If m is very large (2) may be written

$$\frac{v_o}{v_s} = \frac{1}{\frac{1}{m} - \beta} \quad \text{where } \frac{1}{m} \text{ is negligible compared with } \beta.$$

Hence

$$\frac{v_o}{v_s} \approx -\frac{1}{\beta} \dots\dots\dots(3)$$

It follows that if $\frac{1}{m}$ is negligible compared with β , the amplification depends only on β and is practically independent of the amplification produced by the amplifier itself, without feedback. In other words, the output voltage is substantially independent of the characteristics of the valve. The result is obtained only at the expense of amplification.

For example, suppose a pentode amplifier, with an amplification of 500 (without feedback) is used with a feedback circuit having $\beta = +\frac{1}{20}$.

$$\begin{aligned} \text{Then } m = -500 \text{ so that } \frac{v_o}{v_s} &= \frac{-500}{1 - \left(\frac{1}{20}\right)(-500)} \\ &= -\frac{500}{26} \\ &= -19.2 \\ &\approx -\frac{1}{\beta} \end{aligned}$$

Without feedback it would not be possible to maintain a uniform amplification over the whole of the useful portion of the dynamic characteristic, since m would vary considerably in practice. With feedback quite large variations in m may be permitted, the overall amplification remaining approximately equal to $\frac{1}{\beta}$.

In general, provided m is large and a suitable value for β is chosen, amplitude distortion can be made negligible by the use of negative feedback.

It is usual to distinguish between two types of negative feedback, viz, voltage and current feedback. In the former the voltage fed back is proportional to the voltage developed across the amplifier load, whereas in the latter it is proportional to the current delivered to the load. Whilst these produce similar effects for a constant resistive load, they react in different ways to non-resistive or changing loads.

15. Voltage feedback (Parallel Feedback)

The simplest way of achieving voltage feedback is to connect, in parallel with the load R_l of an amplifier, a potentiometer consisting of two resistors of values R_1 and R_2 . These resistances should be large so that the effective value of the load R_l is not modified, and they should be connected in series with a negative supply of steady voltage so that the mean positive potential of the anode is offset and the grid of the amplifier valve suitably biased. Fig. 358 shows the circuit arrangement. The feedback voltage is taken from the junction of the two resistors R_1 and R_2 and is in series with the applied voltage v_s .

The feedback constant β is given by

$$\frac{R_2}{R_1 + R_2}$$

and the voltage fed back will be 180° out-of-phase with the applied voltage if the load R_l is purely resistive. Therefore, if the load is purely resistive and $\frac{1}{m}$ is negligible compared with β , the output

voltage depends entirely on R_1 and R_2 and is quite independent of valve characteristics, supply variations, and the magnitude of R_l . The amplification is substantially independent of the frequency of the sinusoidal components of the input voltage, and negligible phase shift is introduced into the output.

The effect of the feedback connection on the valve performance may be considered by replacing the amplifier, with feedback, by an equivalent valve amplifier without feedback, having the same anode load but with valve constants \bar{G}_m , $\bar{\mu}$, and \bar{R}_a in place of G_m , μ , and R_a . These are defined as follows. If i_a , v_s and v_o are corresponding small changes in anode current, input voltage and anode voltage respectively,

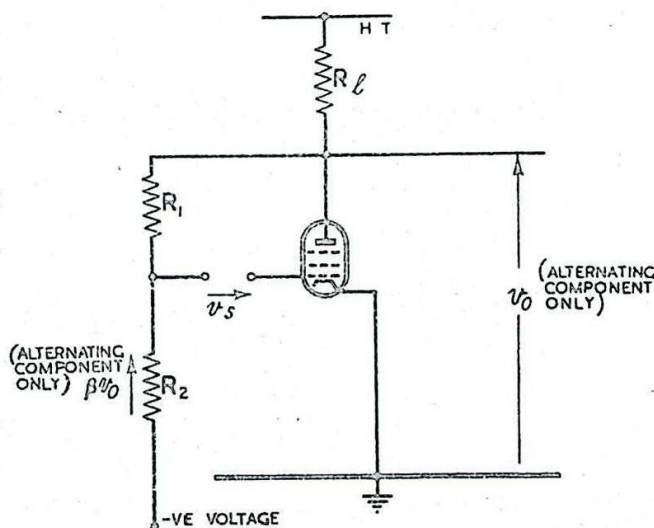


Fig. 358 - Voltage feedback circuit.

$$\bar{G}_m = \frac{i_a}{v_s}, \text{ where anode voltage is constant;}$$

$$\bar{\mu} = - \frac{v_o}{v_s}, \text{ where anode current is constant;}$$

$$\bar{R}_a = \frac{v_o}{i_a}, \text{ where input voltage is constant.}$$

From Fig. 358, the small change βv_o in feedback voltage is in series with v_s so that the small change v_i in grid voltage is given by

$$v_i = v_s + \beta v_o \dots\dots\dots (4)$$

If the anode current is constant, i_a is zero and $v_o = -\mu v_i$

$$= -\mu (v_s + \beta v_o).$$

$$\therefore v_o (1 + \mu\beta) = -\mu v_s$$

$$\therefore \bar{\mu} = -\frac{v_o}{v_s} = \frac{\mu}{1 + \mu\beta} \dots\dots\dots (5)$$

If the anode voltage is constant, v_o is zero and so, therefore, is βv_o . Hence from (4) $v_s = v_i$, and $\bar{G}_m = \frac{i_a}{v_s} = \frac{i_a}{v_i} = G_m \dots\dots (6)$

Since $\mu = G_m R_a$, $\bar{\mu} = \bar{G}_m \bar{R}_a$, and $G_m = \bar{G}_m$, it follows that

$$\frac{\bar{R}_a}{R_a} = \frac{\bar{\mu}}{\mu} = \frac{1}{1 + \mu\beta}$$

$$\text{i.e. } \bar{R}_a = \frac{R_a}{1 + \mu\beta} \dots\dots\dots (7)$$

Hence the amplifier with valve constants G_m , μ , R_a , and with a feedback factor β behaves like an amplifier without feedback but with constants G_m , $\frac{\mu}{1 + \mu\beta}$

and $\frac{R_a}{1 + \mu\beta}$. Since β is

positive for negative feedback $\bar{\mu} < \mu$ and $\bar{R}_a < R_a$,

both being reduced by a factor $\frac{1}{1 + \mu\beta}$.

The equivalent circuit is shown in Fig. 359 (a).

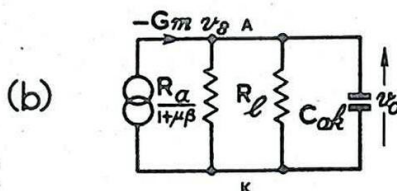
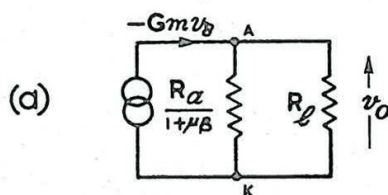


Fig. 359 - Equivalent circuits for voltage feedback amplifier.

In practice it is not possible to make the load purely resistive since there is always some capacitance C_{ak} between the anode and cathode of the amplifier. The equivalent circuit, incorporating this capacitance, is shown in Fig. 359 (b).

The time-constant of this circuit is $\frac{C_{ak} R_l R_a}{R_a + R_l + \mu\beta R_l}$.

If, as for a pentode, $R_a \gg R_l$, this reduces to $\frac{C_{ak} R_l}{1 + G_m \beta R_l}$.

This reduction in the output time-constant increases the fidelity with which sudden changes of input voltage are reproduced at the output. This is illustrated in Fig. 360. For example, with a pentode amplifier having $G_m = 4 \text{ mA/V}$, $R_l = 25,000 \Omega$ and $\beta = \frac{1}{10}$, the time-constant

is reduced by a factor

$$\frac{1}{1 + 0.004 \cdot 0.1 \cdot 25,000} = \frac{1}{11}$$

by the use of voltage feedback. To obtain this result the amplification has been reduced in the same ratio.

In the circuit so far considered the feedback network has not discriminated between different frequencies. It is often required that feedback should occur for signals within some range of frequencies but not for signals outside this range. The commonest instance of this

arises when it is not necessary to consider the amplification of steady voltages. In this case a large blocking condenser can be introduced in series with the feedback potentiometer, as shown in Fig. 361 (a), and the negative bias for the control grid may be provided by a cathode biasing circuit as indicated. This obviates the necessity for providing an additional negative supply voltage to offset the anode voltage.

A circuit which amplifies alternating voltages without feedback but which, owing to negative feedback, provides very little amplification of steady voltages, is illustrated in Fig. 361 (b). The condenser C which is effectively in parallel with R_2 for alternating voltages bypasses current fluctuations so that except for very low frequencies the fraction β of the output voltage which is fed back to the input is small. Also this fraction βv_o is, at high frequencies, in quadrature with v_g . For steady voltages the fraction

$\beta = \frac{R_2}{R_1 + R_2}$ is made large so that the amplification, approximately $\frac{1}{\beta}$, is small.

So far we have ignored the problem of applying the voltage v_g to the input circuit of the voltage feedback amplifier. Since this voltage is applied between terminals neither of which is at a fixed potential relative to earth it cannot be obtained direct from a generator having one terminal earthed. One method of overcoming the difficulty is to use a transformer, as shown in Fig. 362.

The use of a transformer tends to introduce phase and frequency distortion into the input voltage, and its use may be avoided by arranging the circuit as shown in Fig. 363 (a). The equivalent network showing how the voltages are added is given in Fig. 363 (b). R' is the output resistance of the generator, together with any additional series resistance which may be required.

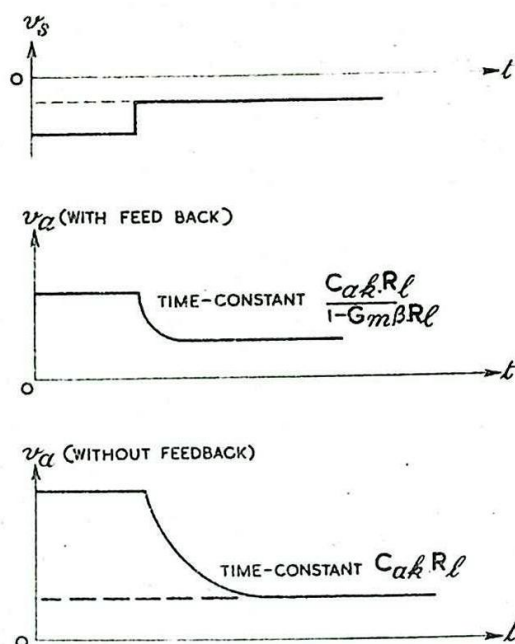


Fig. 360 - Effect of voltage feedback on response of output circuit.

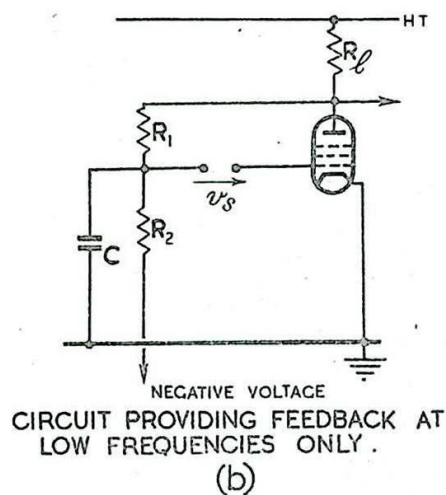
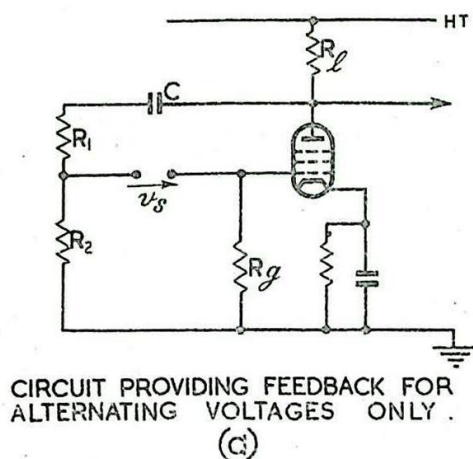


Fig. 361 - Circuit providing feedback (a) for alternating voltages only and (b) at low frequencies only.

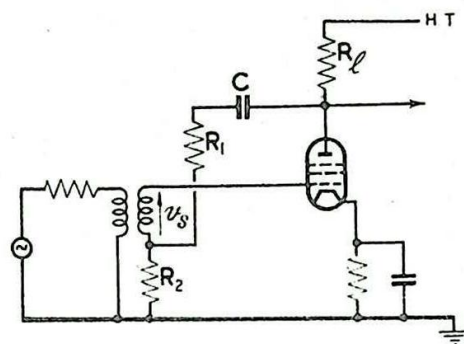


Fig. 362 - Voltage feedback: Transformer input circuit.

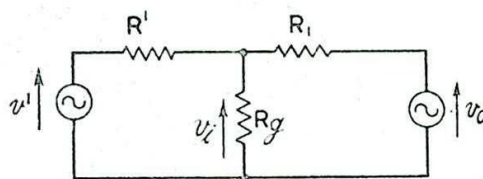
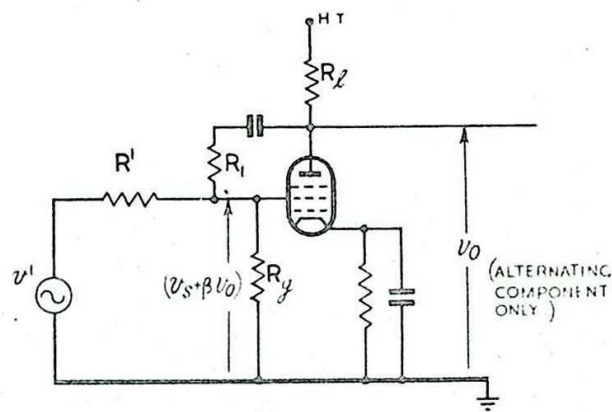


Fig. 363 - Voltage feedback: use of adding network.

The voltage v_i developed across R_g and applied between grid and cathode is given by

$$v_i = \frac{\frac{v'}{R_1} + \frac{v_o}{R_2}}{\frac{1}{R_1} + \frac{1}{R_g} + \frac{1}{R_2}}$$

(See Chap. 1 Sec. 12).

If we write $\frac{1}{R_1} + \frac{1}{R_g} = \frac{1}{R_2}$, this reduces to

$$v_i = \frac{v' R_1 R_2}{R_1 (R_1 + R_2)} + \frac{v_o R_2}{R_1 + R_2},$$

$$= v_s + \beta v_o;$$

where $\beta = \frac{R_2}{R_1 + R_2}$,

and
$$v_s = \frac{v' R_1 R_2}{R' (R_1 + R_2)} .$$

Thus this adding network produces the effect of the arrangement of Fig. 358 without introducing any special problems arising from earthing the generator.

16. Current Feedback (Series Feedback)

To provide negative feedback proportional to the output current it is usual to include a resistor in the cathode circuit of the valve as in Fig. 364 (a). An increase of current through the resistor R_f resulting from a rise of grid voltage causes the cathode voltage also to rise. The effective input voltage developed between the grid and cathode is therefore reduced, i.e., the resistor gives rise to negative feedback.

As for voltage feedback we shall consider the effect of the feedback arrangement in terms of the equivalent amplifier having the same anode load and input and output voltages, but with valve constants \bar{G}_m , $\bar{\mu}$, and \bar{R}_a .

The feedback fraction β , is given by

$$\beta = \frac{R_f}{R_l + R_f} .$$

Let i_a , v_s and v_o represent

corresponding small changes in anode current, input voltage, and anode-cathode voltage respectively. The feedback voltage v_f is then equal to $R_f i_a$, so that the voltage v_i developed between grid and cathode is given by

$$\begin{aligned} v_i &= v_s - v_f \\ &= v_s - R_f i_a \dots\dots\dots (8) \end{aligned}$$

If the anode-cathode voltage is constant, v_o is zero and $i_a = \bar{G}_m v_i$

$$= \bar{G}_m (v_s - R_f i_a) \dots\dots \text{from (8)} .$$

Hence $i_a (1 + \bar{G}_m R_f) = \bar{G}_m v_s$

and $\bar{G}_m = \frac{i_a}{v_s} = \frac{\bar{G}_m}{1 + \bar{G}_m R_f} \dots\dots\dots (9)$

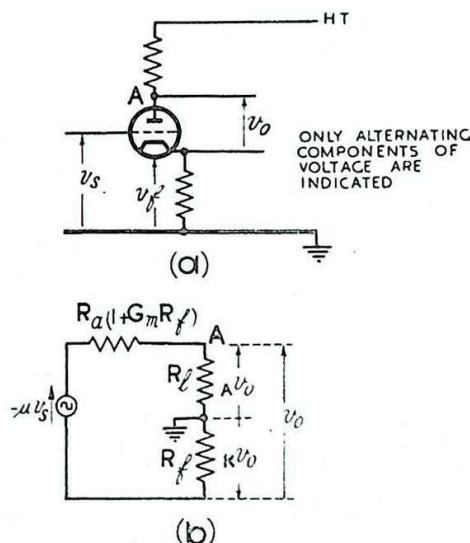


Fig. 364 - Current feedback amplifier.

If the anode current is constant, i_a is zero and so, therefore, is v_f . Hence,

$$\text{from (8), } v_{s.} = v_i \text{ and } \bar{\mu} = -\frac{v_o}{v_s} = -\frac{v_o}{v_i} = \mu \dots\dots\dots (10)$$

Since $\mu = G_m R_a$, $\bar{\mu} = \bar{G}_m \bar{R}_a$, and $\mu = \bar{\mu}$,

$$\frac{\bar{R}_a}{R_a} = \frac{G_m}{\bar{G}_m} = 1 + G_m R_f$$

$$\text{i.e. } \bar{R}_a = R_a (1 + G_m R_f) \dots\dots\dots (11)$$

Hence the amplifier with valve constants G_m , μ and R_a and with a feedback resistor R_f behaves like an amplifier without feedback but with constants $\frac{G_m}{1 + G_m R_f}$, μ , $R_a (1 + G_m R_f)$.

$\bar{G}_m < G_m$ and $\bar{R}_a > R_a$, the one being reduced, and the other increased, by the factor $(1 + G_m R_f)$. The equivalent circuit is shown in Fig. 364 (b).

In this circuit we are interested not only in the voltages developed between anode and cathode, but also in the voltages between anode and earth and between cathode and earth since, in various applications, any of these may be used as output voltages. These will be denoted by v_o , $A v_o$ and $K v_o$ respectively. From Fig. 364(b) we obtain

$$\begin{aligned} v_o &= \frac{R_l + R_f}{R_a (1 + G_m R_f) + R_l + R_f} \cdot (-\mu v_s) \\ &= \frac{R_l + R_f}{R_a + \mu R_f + R_l + R_f} \cdot (-\mu v_s) \\ &= \frac{R_l + R_f}{R_a + R_l + (\mu + 1) R_f} \cdot (-\mu v_s). \end{aligned}$$

Provided μ is large so that μR_f is much greater than the other terms in the denominator, we have

$$\begin{aligned} v_o &\doteq \frac{R_l + R_f}{R_f \mu} (-\mu v_s) \\ &\doteq -\frac{v_s}{\beta}. \end{aligned}$$

$$\text{Hence } m \doteq \frac{v_o}{v_s} \doteq -\frac{1}{\beta} \dots\dots\dots (12)$$

$$\begin{aligned} \text{Since } A v_o &= \frac{R_l}{R_l + R_f} v_o \\ &= (1 - \beta) v_o \dots\dots\dots (13) \end{aligned}$$

we deduce that

$$\frac{A^V_o}{v_s} = - \frac{(1-\beta)}{\beta}$$

Similarly, $K^V_o = -\beta v_o$

$$= -\beta \left(- \frac{v_s}{\beta} \right)$$

so that

$$\frac{K^V_o}{v_s} = 1 \dots\dots\dots (14)$$

In the case of current feedback the increase in the output resistance of the valve, and hence in the output resistance of the amplifier, increases the effect of stray capacitance in parallel with the anode load. The increase in the output time-constant can be offset by the addition of a suitable capacitance in parallel with R_f , thereby increasing the high-frequency response.

Where the amplifier has a choke load instead of a resistor, as, for instance, in a circuit for producing magnetic deflection currents for a CRT, the increase in the output resistance of the valve due to current feedback is an advantage since it reduces the time-constant of the output circuit.

In the above consideration of current feedback we have assumed that the valve is a triode. If a pentode is used two points should be noted. The screen grid should normally be decoupled to cathode and not to earth, otherwise the effective screen potential varies and the amplification is reduced. Also the current through the cathode feedback resistor consists of screen, as well as anode, current. If the variations in screen current are neglected the effect will be the same as if the resistor R_f were assumed to have a value larger than it actually has.

In all cases of current feedback difficulties arise if the cathode potential is allowed to differ substantially from that of earth, with the possibility of insulation breakdown between cathode and heaters. This difficulty may be overcome by using special heater windings carefully insulated from earth.

17. Bias for Feedback Amplifiers

Where no blocking condenser is inserted in the feedback network, an amplifier employing negative voltage feedback normally requires a large negative bias to offset the positive potential at the anode. This may be provided either by a separate negative supply or by means of a cathode bias resistor, suitably decoupled.

If a blocking condenser is used the bias required is no different from that needed by an ordinary amplifier.

The steady voltage developed by the direct current of the valve through the cathode feedback resistor automatically provides a bias voltage for a current feedback amplifier. This bias may be just sufficient to ensure that the valve is operated over the desired portion of its dynamic characteristic, but generally is either too large or too small.

Fig. 365 (a) illustrates the case in which the mean feedback voltage $R_f \bar{i}_a$, where \bar{i}_a is the mean anode current, is just sufficient to provide the correct bias.

Fig. 365 (b) shows the use of an additional cathode bias resistor R_k suitably decoupled, for the case in which $R_f \bar{i}_a$ is not large enough to provide the correct bias.

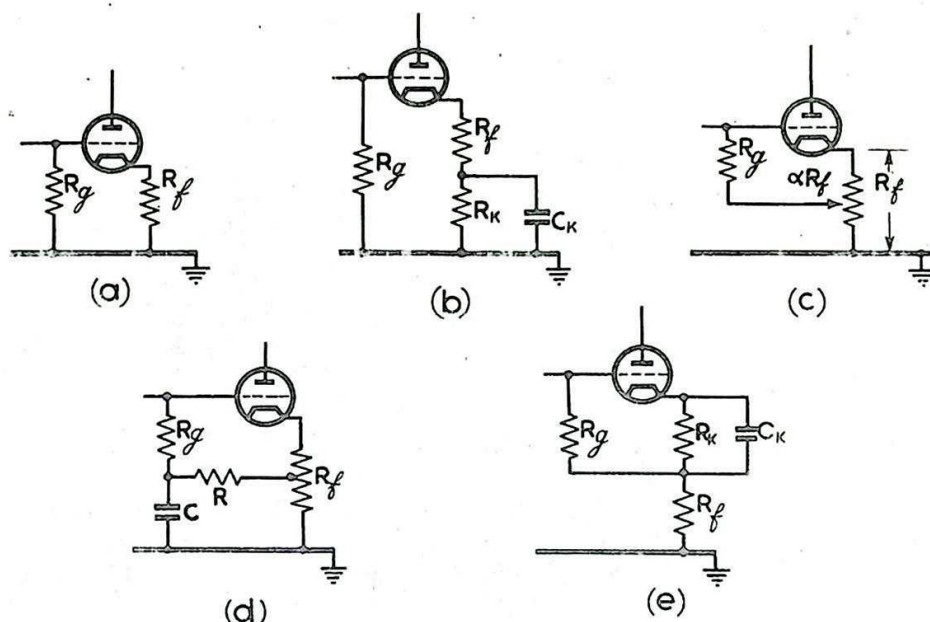


Fig. 365 - Current Feedback: biasing arrangements.

In the arrangement of Fig. 365 (c) $R_f \bar{i}_a$ provides more bias than is necessary. A point on R_f is chosen at which the mean potential, $R_f \bar{i}_a$ below cathode potential, has the correct value, and the lower end of the grid leak is connected to this point. If the output impedance of the generator feeding the grid circuit is large enough to be comparable with R_g , additional feedback to the grid circuit will occur due to this arrangement. This effect may be substantially eliminated by the insertion of a decoupling condenser C connected to the lower end of the grid leak (d). The additional resistor R is necessary to provide the steady voltage connection but must be of sufficiently large value to prevent the lower portion of R_f from being bypassed.

The network of Fig. 365 (e) may be used to provide any degree of negative bias irrespective of the size of the feedback resistor. This method is not economical if a large bias is required.

Sometimes voltage and current feedback are used in the same valve circuit. By this method the output voltage and current can both be made substantially independent of valve characteristics. No additional biasing problems are introduced if the voltage feedback network contains a blocking condenser.

Fig. 366 (a) illustrates the simple case in which voltage and current feedback both operate for steady as well as alternating voltages. The voltage $R_f \bar{i}_a$ is just sufficient to offset the positive voltage at the grid due to the feedback potentiometer, and to provide the correct bias.

In the circuit of Fig. 366 (b) feedback occurs for very low frequencies only. The condenser C ensures that alternating input voltages are effectively developed between grid and cathode so that no appreciable feedback occurs. R_g is connected to a suitable biasing point on R_f .

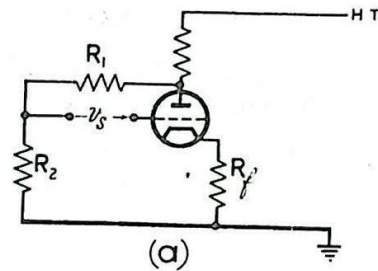
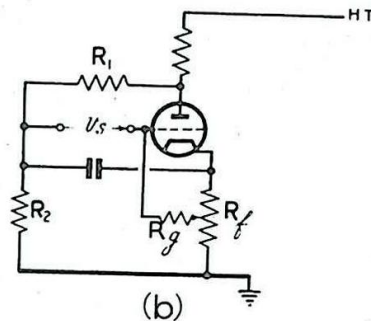


Fig. 366 - Biasing arrangements for circuits using combined current and voltage feedback.



THE CATHODE FOLLOWER

18. The Properties of a Cathode Follower

In this type of current feedback amplifier the output is taken from the cathode circuit and there is no anode load. The output is the voltage developed across the cathode load by the fluctuating component of the cathode current, which in a triode is the same as that of the anode current. The presence of negative current feedback causes the cathode follower to have the following main properties.

- (1) A voltage amplification which may be approximately equal to, but is always less than, unity.
- (2) A low output resistance.
- (3) A small input capacitance.

Fig. 367 (a) shows a typical cathode follower circuit. A resistor may be included in the anode circuit to ensure that there is the correct steady potential on the anode for normal operating conditions, but this resistor is usually decoupled to earth by means of a large condenser so that there are no fluctuations of anode potential.

The response of the cathode follower circuit to alternating voltages may be derived from the results of Sec. 16, or may be obtained independently in the following manner.

Fig. 367 (b) shows an equivalent circuit for the arrangement of Fig. 367 (a). From (a) we obtain $v_i = v_g - v_o$, whilst from (b)

$$v_o = - \frac{R_f}{R_f + R_a} \cdot (-\mu v_i)$$

$$= \frac{R_f}{R_f + R_a} \cdot \mu (v_g - v_o).$$

$$\text{i.e. } v_o (R_f + R_a) = \mu R_f v_s - \mu R_f v_o.$$

$$\text{Hence } v_o = \frac{\mu R_f v_s}{R_a + \mu R_f + R_f}.$$

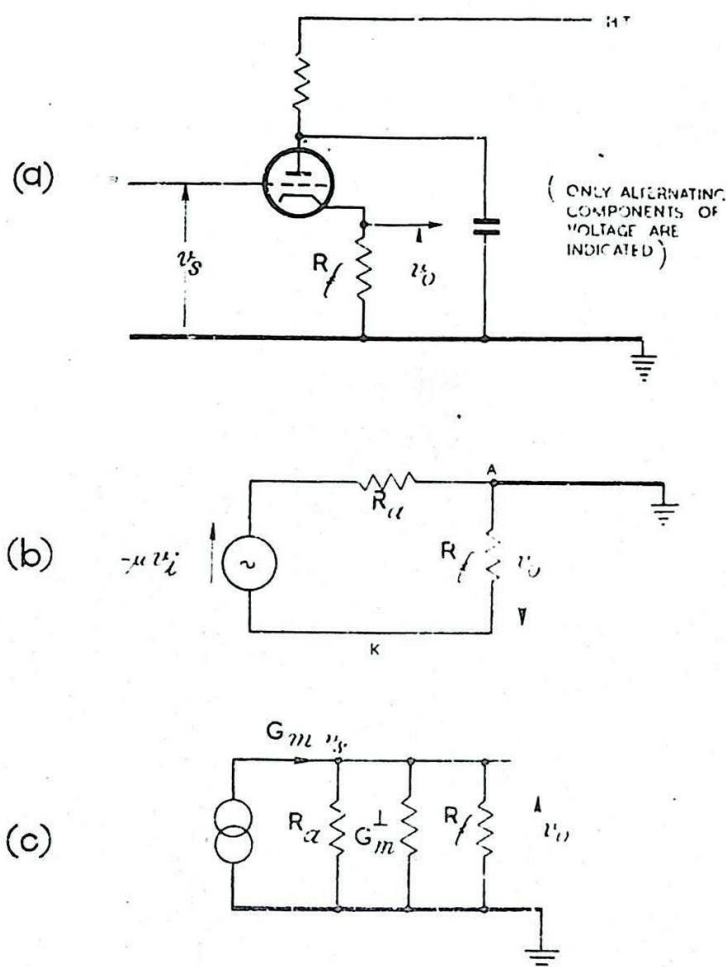


Fig. 367 - Cathode follower.

Writing $\mu = G_m R_a$, and dividing numerator and denominator by $R_a R_f$, we obtain

$$v_o = \frac{G_m v_s}{\frac{1}{R_f} + G_m + \frac{1}{R_a}} \dots\dots\dots (15)$$

If R_f and R_a are both much greater than $1/G_m$, $v_o \doteq v_s$.

Equation (15) shows that the equivalent circuit may be drawn as indicated in Fig. 367 (c). Since G_m is normally a few milliamps per volt whereas R_a is several thousand ohms the output resistance of the cathode follower presented to the load R_f is approximately $\frac{1}{G_m}$, with a value equal to a few hundred ohms. This is very much less than the output resistance of a conventional amplifier, and is of a value suitable for matching to a cable or to an artificial line.

Since the output resistance of the circuit is small, the time-constant due to stray capacitance C in parallel with the load is also small so that there is less likelihood of distortion from this cause than in a normal amplifier. If, however, the rate of change of input voltage is rapid enough, the presence of capacitance in the output circuit prevents the cathode voltage from following that of the grid instantaneously. This may cause grid current to flow if the input is positive-going, or may cut off the valve current if it is negative-going. In the latter case the output resistance of the valve becomes infinite and the output time-constant of the circuit becomes CR_p .

As an example, consider a triode valve operating as a cathode follower under the following conditions.

Slope of dynamic characteristic	= 2 mA/V
Maximum value of grid-cathode voltage	= 0 volts
Minimum value of grid-cathode voltage	= -10 volts
Maximum value of anode current	= 22 mA
Cathode feedback resistor	= 10 k Ω

Assuming linearity, the minimum value of the anode current is $22 - 2 \cdot 10 \text{ mA} = 2 \text{ mA}$. For this value, the grid-cathode voltage is -10 volts and the cathode-earth voltage is $10,000 \cdot 0 \cdot 002 = 20 \text{V}$.

When the grid-cathode voltage is zero, the anode current is 22 mA and the cathode-earth voltage is $10,000 \cdot 0 \cdot 022 = 220 \text{V}$.

Hence the output voltage varies from 20 to 220 volts while the input varies from 10 to 220 volts.

The mean grid-cathode voltage is -5 volts, and if the bias is obtained by connecting the grid leak to a point on the cathode feedback resistor, the connection should be made at a point approximately 400 ohms from the cathode. Since the mean anode current is $\frac{22 + 2}{2} = 12 \text{ mA}$, the bias is then $-0 \cdot 012 \cdot 400 \approx -5 \text{V}$.

If in this example the grid voltage were allowed to exceed 220 volts grid current would flow even if the output time-constant were zero. If the grid voltage were allowed to fall below 10 volts the valve would operate over the lower bend of the dynamic characteristic and the anode current might possibly be cut off.

The relative unimportance of Miller effect in a cathode follower, causing it to present a much smaller input capacitance than that of a similar valve in an ordinary amplifier circuit, will now be considered. The relevant interelectrode capacitances are shown in Fig. 368 (a) and the equivalent circuit is given in Fig. 368 (b). Since the anode is at earth potential so far as alternating voltages are concerned, C_{ga} is in effect across the input terminals. By an analysis similar to that adopted in Sec. 7, for the circuit of Fig. 338, we obtain

$$y_i = \omega C_{gk} |m| \sin \theta + j \omega \left[C_{ga} + C_L \cdot 1 - |m| \cos \theta \right]$$

If the effective load of the cathode follower is purely resistive the cathode and grid voltages are in phase so that $\theta = 0$. The input admittance is then a capacitive susceptance of value

$$\omega [C_{ga} + C_{gk} (1 - m)]$$

Consequently the effective input capacitance of the circuit is equal to $C_{ga} + C_{gk} (1 - m)$.

Since $m \approx 1$, this normally reduces to C_{ga} . Even if the load impedance is not purely resistive, the maximum value which the input capacitance can acquire is $C_{ga} + C_{gk}$.

The input conductance, when the load is not purely resistive is given by

$$\omega C_{gk} |\sin \theta|$$

If the effective load is inductive, $\sin \theta$ is positive and the input conductance has a damping effect on the input circuit. If the effective load is capacitive, $\sin \theta$ is negative and the circuit is regenerative; this condition may give rise to continuous oscillations.

19. Applications of a Cathode Follower

Because the output resistance of a cathode follower is normally much lower than that of a conventional amplifier, the cathode follower is useful for feeding voltages into a relatively low impedance load such as that presented at video-frequencies by the input capacitance of a coaxial cable. Such a cable is often used for carrying voltage pulses from one part of a radar equipment to another.

A cathode follower circuit is also useful for driving a non-linear circuit, to minimise the distortion due to this non-linearity. A typical example of a non-linear circuit is the input circuit of an amplifier in which grid current flows for part of the time. The input resistance, i.e., the resistance between grid and cathode of an amplifier, is usually of a high value, but if grid current flows, is considerably reduced. If an amplifier is fed from the output of a cathode follower the input resistance of the amplifier shunts the output resistance of the cathode follower (constant current circuit, Fig. 367 (c)). However, since the output resistance of the cathode follower is usually very small, the effect when the input resistance of the amplifier changes from a high to a lower value is not large. Consequently the amount of distortion introduced by non-linearity of the input circuit of the amplifier is small. An additional advantage, under the circumstances described above, is that the cathode of the cathode follower can often be directly connected to the grid of the amplifier, since the steady voltage at the cathode is usually of a low value. If this arrangement is possible, the use of a coupling condenser

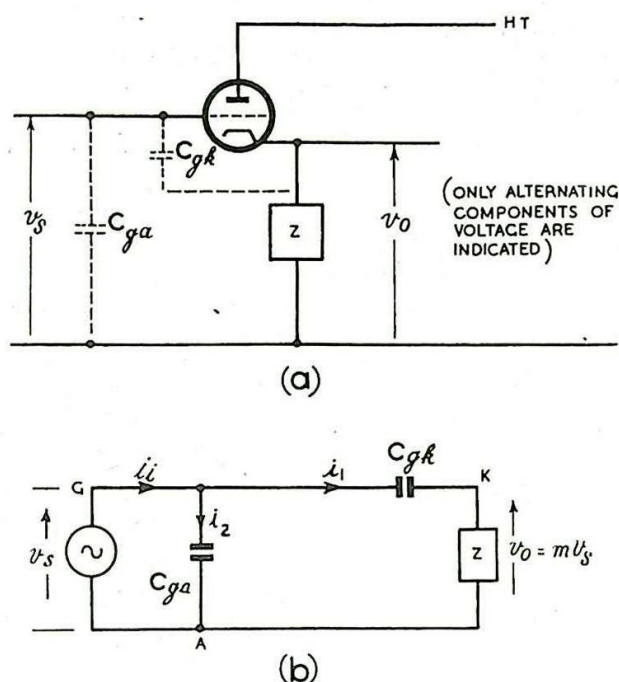


Fig. 368 - Miller effect in a cathode follower.

and grid leak resistor, usually essential in the input circuit of an amplifier, is avoided. If there is no coupling condenser or grid leak resistor, the possibly undesirable effects of slide-back bias due to grid current are avoided. Another common instance of the use of a cathode follower to avoid distortion due to a non-linear load arises in feeding the deflector plates of a CRT when deflector plate current flows.

If it is required to feed the output of an amplifier into a load circuit which has a large input capacitance a cathode follower is particularly useful. The output from the amplifier is taken to the input circuit of the cathode follower, and the output of the cathode follower is fed to the load. The small input capacitance of the cathode follower is in shunt with the output of the amplifier so that the time-constant is small and the potential of the anode of the amplifier (and of the grid of the cathode follower) can change rapidly.

The output resistance of the cathode follower is in shunt with the input capacitance of the load, so that the output time-constant also is small and the load voltage can change rapidly. A cathode follower used in this way may be regarded as an impedance transformer, i.e., the input circuit presents a high impedance whilst the output circuit presents a low impedance.

Since it can be arranged that the cathode potential of a cathode follower rises to almost the same value as the grid potential, it is possible to connect the grid of the cathode follower directly to a point at a high positive potential, without causing grid current to flow. This means that the use of a coupling condenser and grid leak resistor may be avoided. A cathode follower may, therefore, be connected across a circuit, the operation of which would be upset by a resistive loading. For example, suppose that a condenser is charged through a constant-current device so that the voltage across the condenser rises linearly with time. If there is resistance in parallel with the condenser this linearity is affected. A direct-coupled cathode follower may be used to transfer the linear voltage change across the condenser to another part of the circuit, since it does not appreciably shunt the condenser with resistance; (Fig. 369).

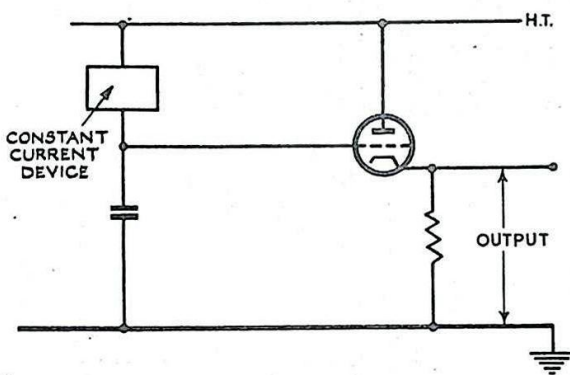


Fig. 369 - Use of direct-coupled cathode follower to minimise loading on a time-base generator.

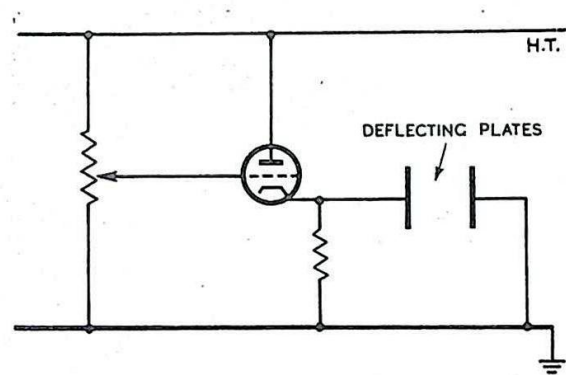


Fig. 370 - Use of a direct-coupled cathode follower to minimise loading on a potentiometer.

It is often necessary to arrange that the potential on the slider of a potentiometer (Fig. 370) varies linearly with the position of the slider. This cannot happen if there is shunt resistance between the slider and one or other of the ends of the potentiometer. If,

however, the slider is directly coupled to the grid of a cathode follower, the voltage of the cathode of this valve varies linearly with the position of the slider on the potentiometer. In Fig. 370 the cathode of the cathode follower is shown to be directly connected to a deflecting plate of a cathode ray tube. This arrangement makes it possible to apply to the deflector plate a shift potential which varies linearly with the position of the slider on the potentiometer.

20. THE CATHODE INPUT (OR GROUNDED GRID) AMPLIFIER

The Cathode Input Amplifier is complementary to the cathode follower and is, therefore, considered at this point. An important feature of this circuit is its low input resistance, which makes it suitable for terminating a long cable to which it may be matched. The same property makes the circuit useful as an RF amplifier, since it helps to prevent regeneration. The circuit has the comparatively high output resistance of a normal amplifier.

The circuit arrangement is shown in Fig. 371 (a).

This circuit differs from that of a normal amplifier in that the negative HT supply lead is connected to the grid of the valve instead of to the cathode, so that the anode current flows through the input circuit. The grid is normally earthed, and the input voltage is applied to the cathode. The anode and cathode voltages are in phase, since when the cathode voltage is reduced the current through the valve increases and the anode voltage falls also.

Fig. 371 (b) shows an equivalent circuit. The resultant EMF in the circuit is $-(\mu + 1) v_i$ so that the output voltage developed across R_f is given by

$$v_o = - \frac{R_f}{R_a + R_f} (\mu + 1) v_i$$

and the amplification by

$$m = \frac{v_o}{v_i} = - \frac{(\mu + 1) R_f}{R_a + R_f}.$$

The amplification is slightly greater than that due to a normal amplifier using the same valve and anode load.

The alternating component i_a of the anode current is given

$$i_a = \frac{(\mu + 1) v_i}{R_a + R_f}.$$

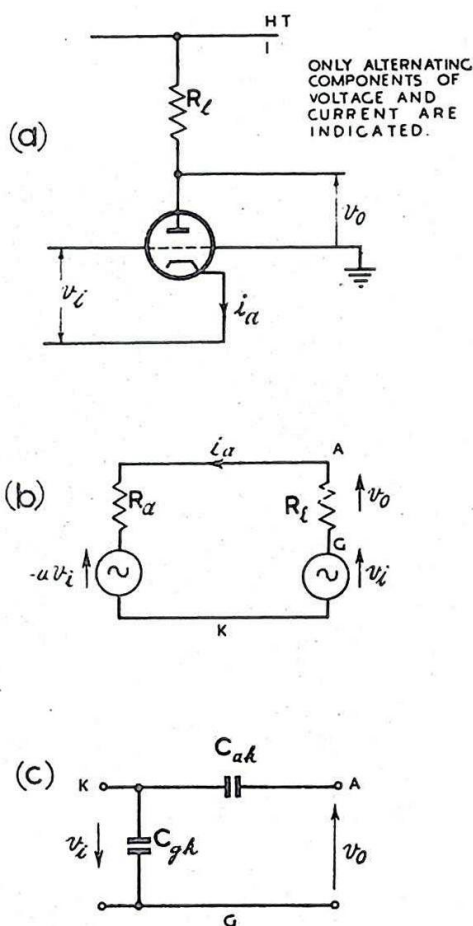


Fig. 371 - Cathode input amplifier.

Since this current flows through the input circuit, in phase with v_i , the input resistance R_i of the amplifier is given by

$$R_i = \frac{v_i}{i_a} = \frac{R_a + R_l}{\mu + 1}.$$

If $R_a \gg R_l$ and $\mu \gg 1$, this reduces to

$$R_i \approx \frac{R_a}{\mu} = \frac{1}{G_m}.$$

The input circuit of a cathode input amplifier contributes towards the output power, the fraction supplied being $\frac{1}{\mu}$ of the power derived from the HT supply. This makes the circuit particularly useful as a power amplifier.

Miller effect is of little importance since it occurs through feedback via C_{ak} and this capacitance is usually very small. (Fig. 371 (c)). This fact enables triodes to be used without neutralising. Since the input resistance of the amplifier in any case is small, the current fed back from the output circuit via C_{ak} must be very large before regeneration occurs.

PARAPHASE AMPLIFIERS

21. General

The purpose of a Paraphase Amplifier is to amplify an unbalanced input voltage so as to produce a balanced output (See Chap. 3 Sec 1). This could be accomplished by means of a transformer with the primary winding earthed at one end and the secondary winding earthed at its electrical centre, as illustrated in Fig. 372. Since the design of transformers for undistorted pulse amplification is complicated the employment of a paraphase amplifier is normally preferable. Single-valve and two-valve amplifiers are described.

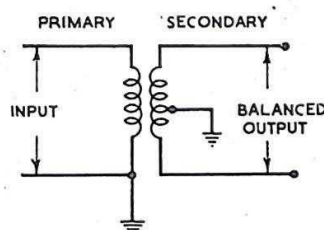


Fig. 372 - Transformer used for paraphase operation.

22. Single-Valve Paraphase Amplifiers

The simplest type of paraphase amplifier employs a single valve, the input for which is a fraction $\frac{1}{|\mu|}$ of the available applied

voltage. This fraction is amplified $|m|$ times, and inverted, so that the output voltage is equal to the applied voltage and of opposite polarity. Fig. 373 shows the circuit arrangement.

The potentiometer formed by R_1 and R_2

provides the fraction

$$\frac{1}{|m|} = \frac{R_2}{R_1 + R_2} \text{ of the}$$

applied voltage v_s .

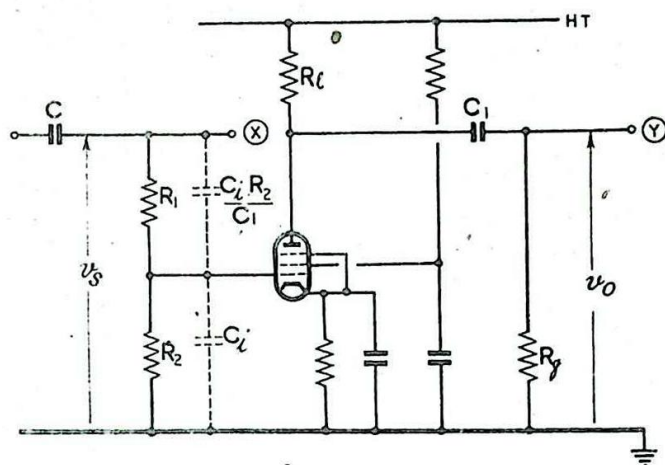


Fig. 373 - Simple one-valve paraphase amplifier.

The condensers C and C_1 eliminate the steady components applied and output voltages respectively, so that both v_s and v_o have a mean value zero and $v_o = -v_s$.

Distortion of the input voltage due to input capacitance C_1 may be minimised by adding capacitance between the grid and the point X (Fig. 373) so that the time-constants of both portions of the potentiometer are equal.

One disadvantage of this circuit is that any variation in the valve characteristics or supply voltages affects the valve amplification and unbalances the outputs. The amplifier may be made linear and stable by introducing negative feedback. The simplest way of achieving this is by the introduction of a resistor R_f in the cathode circuit (Fig. 374). If a pentode is used in this circuit, the inequality between corresponding changes of anode and cathode currents makes the balance dependent on the ratio between these currents being constant. A triode cannot always be used, in place of the pentode, owing to the larger input capacitance due to Miller effect. Consequently it is advantageous to use voltage feedback. In the circuit shown in Fig. 375, the voltage feedback is introduced by means of a potentiometer, consisting of the resistors R_1 and R_g . The fraction of the input voltage which is actually applied to the valve is determined by the relative values of the resistors R' and R_g . It is found that provided the ratio $\frac{R_1}{R_g}$ is small and the amplification

of the valve circuit is large then paraphase amplification is obtained when $R' = R_1$.

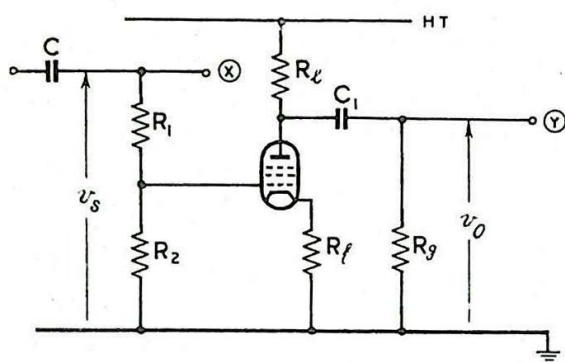


Fig. 374 - Paraphase amplifier with current feedback.

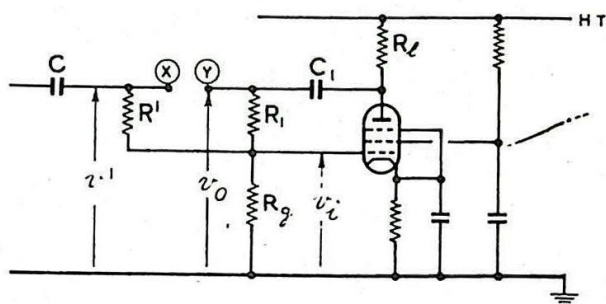


Fig. 375 - Paraphase amplifier using voltage feedback.

*** Analysis

This circuit is essentially the same as that of Fig. 363(a). For that circuit we obtained in Sec. 15 the results

$$v_i = v_s + \beta v_o$$

where
$$\beta = \frac{R_2}{R_1 + R_2},$$

$$v_s = \frac{v' \cdot \frac{R_1}{R'} \cdot R_2}{R' (R_1 + R_2)} = v' \frac{R_1}{R'} \beta,$$

and
$$R_2 = \frac{R_g R'}{R_g + R'}.$$

In this case we want $v_o = -v'.$

Then, since $v_o = m v_i$, where m is the amplification of the valve (without feedback),

$$v_o = m (v_s - \beta v')$$

$$\therefore -v' = m \left(\frac{v' R_1 \beta}{R'} - \beta v' \right)$$

$$= m v' \beta \left(\frac{R_1}{R'} - 1 \right).$$

Hence $\frac{R_1}{R'} - 1 = -\frac{1}{m\beta}$, where m is negative.

Provided $-m\beta$ is very large, $\frac{R_1}{R'} \approx 1$ for the output to be correctly balanced. β is large provided R_1 and R' are small compared with R_g , so that the condition $R_1 \approx R'$ holds if the amplification is large and $\frac{R_1}{R_g}$ is small.

An alternative method of obtaining a paraphase output, using a single valve, depends on the fact that in a valve with purely resistive anode and cathode loads the anode and cathode potentials vary in opposite phase. In the circuit shown in Fig. 376 the anode and cathode loads are equal in value. In the absence of grid current flow, changes in valve current due to changes of input voltage produce equal voltage changes across the two equal loads. A paraphase output is, therefore, developed between the anode and cathode of the valve.

The circuit is a particular case of an amplifier with current negative feedback, so that the amplification characteristics of the circuit are linear and stable.

Since the potential change at the cathode cannot be greater than the change at the input, and is equal and opposite to the potential change at the anode, the overall amplification of the stage is not greater than 2.

The output resistance of the valve when viewed from the anode is different from that viewed from the cathode. When stray capacitances are considered the time-constants of anode and cathode circuits are generally different so that when the input signal consists of pulses the distortion is different in the two outputs.

Fig. 376 shows the valve as a triode. There is no serious disadvantage in using a triode, since, though Miller effect is present, that part of the input admittance, which is due to the grid-anode interelectrode capacitance, is not large since the amplification is small, so that, like the cathode follower, the circuit has a small input capacitance. If a pentode is used the resistance of the cathode load has to be made somewhat smaller than that of the anode load since the change of cathode current for a given change of applied voltage is somewhat larger than the corresponding change of anode current.

*** Analysis

The results (12), (13) and (14) of Sec. 16 give approximate values for the amplification of the circuit of Fig. 376 if R_l is made equal to R_f . This makes $\beta = \frac{1}{2}$ so that $m \approx -2$. The analysis of Sec. 16 gives the exact value as

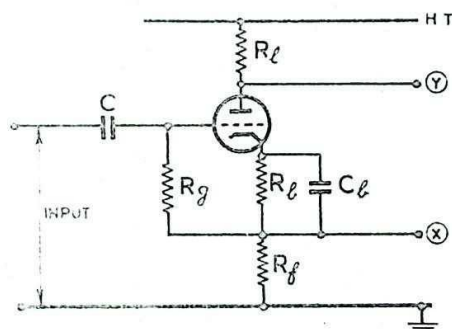


Fig. 376 - Paraphase amplifier using current feedback.

$$m = -\mu \cdot \frac{2R_l}{R_a + (\mu + 2) R_l}, \text{ so that the amplification}$$

produced at each output is

$$\frac{\mu R_l}{R_a + (\mu + 2) R_l}.$$

23. Two-Valve Paraphase Amplifiers

In the single-valve paraphase amplifiers described in Sec. 22, the valve acts as a phase reverser, but in no case does the magnitude of each output voltage exceed that of the applied voltage. If a greater output is required in paraphase two valves may be used as described below.

The most obvious manner in which two valves may be used is by introducing, before the circuit of Fig. 373, a single stage of amplification. This arrangement gives the circuit shown in Fig. 377. It is normal, though not essential, for the values of the anode load and of the mutual conductance of the second valve to be the same as those of the first valve. Then, as shown in Sec. 22 in reference to Fig. 373, the condition for a paraphase output is

$$\frac{m \cdot R_2}{R_1 + R_2} = -1,$$

where m is the amplification of each valve circuit. Each of the amplifying stages introduces distortion as a result of the non-linearity of the valve characteristics and as a result of interelectrode and stray capacitances. However, though the first stage may give as its output a distorted version of the input voltage, it is distortion introduced by the second stage which prevents perfect paraphase. The distortion introduced and the dependence of the circuit on valve characteristics can, of course, be reduced by using negative feedback in each of the amplifying stages. A condenser may also be introduced in parallel with resistor R_1 in order to avoid the distortion which would otherwise be introduced by the input capacitance of the second valve, as explained in Sec. 22.

An amplifying stage may also be introduced before the circuit of Fig. 375. This type of circuit is often known as the Floating Paraphase Amplifier, and is shown in Fig. 378. This first stage can be considered as a normal amplifier, though its amplification is modified by the network of resistors R' , R_g and R_l , the input resistance of which is in parallel with the load resistor. This effect is negligible provided the input resistance of the network is large compared with the load resistance. In the second amplifier stage, the variations of potential of the point O provide the input voltage to the second valve, and these variations depend on the changes in potential at the anodes of the two valves, which are antiphase.

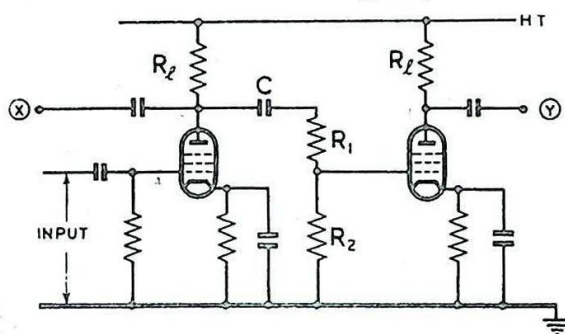


Fig. 377 - Simple two-valve paraphase amplifier.

Usually the resistor R' is made equal to the resistor R_1 and it can be seen that perfect paraphase is then impossible, since equal and opposite changes of potential at the two anodes produce zero variation of potential of the point O, and, therefore, no input voltage to the second valve. In this arrangement the output from the second valve is less than that from the first valve by just that amount necessary to produce at the point O variations which, after

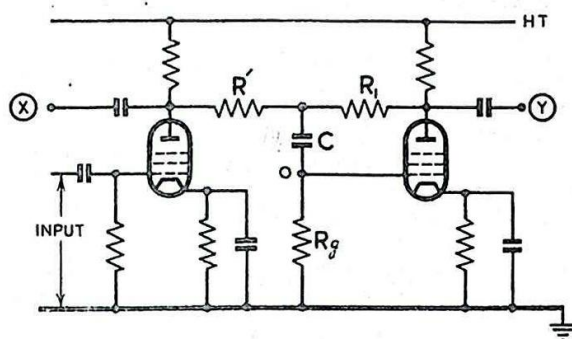


Fig. 378 - Floating paraphase amplifier.

amplification, give the output from the second valve. As in the case discussed in Sec. 22, provided $\frac{R_1}{R_g}$ is small, and the amplification of

the second stage is large, the changes in potential at the two anodes can be regarded as equal in amplitude. There is negative voltage feedback operating on the second stage so that the operation of the second stage is stable, more or less independent of valve characteristics, and produces little distortion.

The conditions for exact paraphase working are the same as those given in Sec. 22 for the circuit of Fig. 375.

A further method of obtaining a paraphase output voltage is provided by the Cathode Inversion Circuit (or cathode-coupled paraphase amplifier) shown in Fig. 379 (a). It consists of two valves coupled together by the common cathode resistor R_5 , which is usually of several thousand ohms and is connected to a large constant negative potential $-V$. The values of the resistor R_5 and the potential $-V$ (about -200 volts) are adjusted so that in the absence of any applied voltage the valves are both biased in Class A. The anode load resistors R_1 and R_2 are usually made equal. The input voltage is applied between the grid of the first valve and earth, and the potential of the other grid is maintained constant, in this case at earth potential.

Initially, if there is no applied voltage, both valves are biased by an amount equal to the difference between $-V$ and the voltage developed across R_5 by the flow of the combined valve currents.

Consider an actual circuit with $R_5 = 5 \text{ k}\Omega$, and the slope of the dynamic characteristic of each valve equal to 1 mA/volt. An increase of one mA through R_5 raises the potential of K by 5 volts, decreasing the cathode current of valve 2 by 5 mA. This must be accompanied by an increase of 6 mA in valve 1, due to an increase of 6 volts between its grid and cathode, and therefore of 11 volts between its grid and earth. Thus any input voltage is distributed between the two valves in the ratio 6 : 5, so that with equal anode loads the circuit is unbalanced in this ratio. The larger the value of R_5 the more nearly balanced are the output voltages; at the same time the voltage $-V$ must be made more negative so that the mean valve current is unchanged.

Since the grid base of either valve is unlikely to be more than about 15 volts, the change of potential at K is small compared with the voltage across R_5 (100V to 200V). Hence the current through

R_5 is approximately constant so that increases in the current of one valve are accompanied by approximately equal decreases in the current of the other.

Instead of the lower end of R_5 being connected to a negative potential, bias may be provided for the grids by connecting them to a source of constant positive potential. This is usually more convenient, but considerably diminishes the effective HT voltage, developed between the positive HT line and cathode.

Since approximately half the input voltage is developed between the grid and cathode of each valve the overall amplification of the circuit is approximately the same as that for a single valve of the same type with the same anode load ($R_1 = R_2$).

The equivalent circuit for Fig. 379 (a) is shown at (b). In this circuit $v_s = 1v_i - 2v_i$, where v_s is the applied voltage and $1v_i$, $2v_i$ the input voltages between the grid and cathode of valves 1 and 2 respectively. Similarly $1v_o$ and $2v_o$ are the alternating components of the output voltages measured between the anodes and earth.

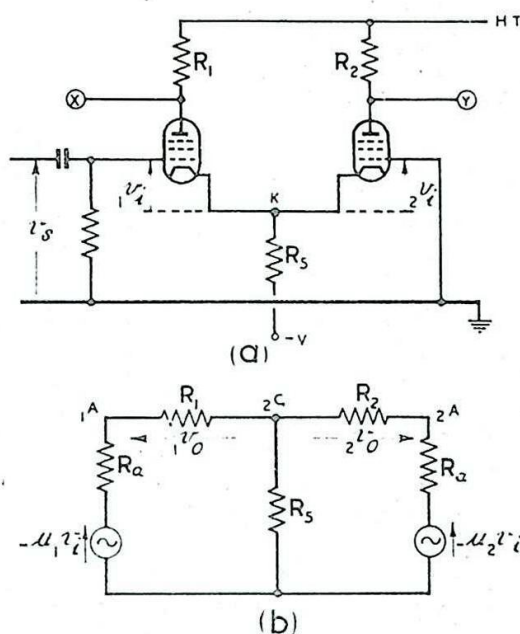


Fig. 379 - Cathode-coupled paraphase amplifier.

Analysis shows that the condition for true paraphase working, assuming that the valves have identical characteristics, is

$$(\mu + 1) \frac{(R_2 - R_1)}{R_a + R_2} = \frac{R_1}{R_5}$$

Provided $\mu \gg 1$ and $R_a \gg R_2$, this gives

$$R_2 - R_1 \doteq \frac{R_1}{G_m R_5}$$

Hence good paraphase working with $R_1 = R_2$ is possible only if $G_m R_5 \gg 1$.

On putting $R_1 = R_2$ in the expressions for $1v_o$ and $2v_o$ it may be shown that the overall amplification $m = \frac{1v_o - 2v_o}{v_s} = -\frac{\mu R_1}{R_a + R_1}$

as it would be for a single valve with the same characteristics and the same anode load as used for either valve in the paraphase circuit.

The output voltages depend on the valve characteristics, no stabilising effect being introduced by the cathode load R_5 . Such an

effect can be obtained by the use of negative feedback separately for each of the valves. This may be achieved, for example, by introducing resistors R_3 and R_4 as shown in Fig. 380. By applying a star-delta transformation to the network formed by R_3 , R_4 and R_5 the equivalent network formed by R_6 , R_7 and R_8 is obtained (Fig. 381).

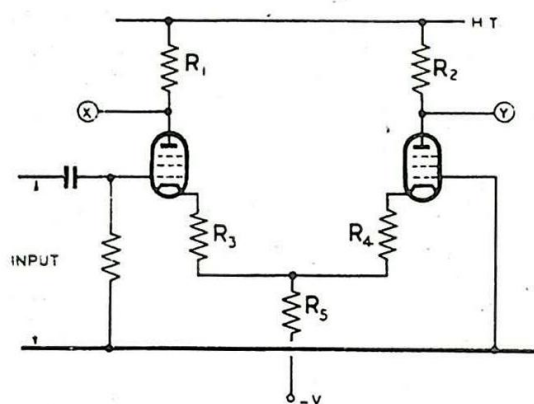


Fig. 380 - Cathode-coupled paraphase amplifier with current feedback (Y-network).

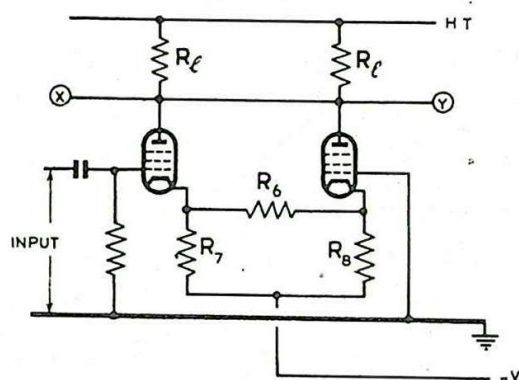


Fig. 381 - Cathode-coupled paraphase amplifier with current feedback (Δ -network).

The relations between these networks may be written :-

$$R_6 = \frac{R_3 R_4 + R_4 R_5 + R_5 R_3}{R_5},$$

$$R_7 = \frac{R_3 R_4 + R_4 R_5 + R_5 R_3}{R_4},$$

$$\text{and } R_8 = \frac{R_3 R_4 + R_4 R_5 + R_5 R_3}{R_3}.$$

This is a practical variant of the cathode inversion circuit.

In practice the presence of stray capacitance across the output terminals, and from each anode to earth, affects the rapidity of response of the amplifier to rapidly changing input voltages. The output impedance of the circuit is comparable in magnitude with that of a normal amplifier using a valve and load similar to those of one amplifier of the paraphase circuit, so that the output time-constant is not substantially altered by the paraphase connection. This is not necessarily true if negative feedback is used.

The cathode inversion circuit may be used to supply a paraphase voltage change to a pair of deflecting plates of a cathode ray tube, and the output of the circuit can be taken direct to the plates without the inclusion of coupling condensers. In this case paraphase shift potentials can be obtained and varied in amplitude by alteration of the steady potential of the grid of the second valve, since this results in amplitude changes of steady potential in opposite directions at the anodes of the two valves. The use of a separate "shift" network is avoided. If a variable resistor is included in the cathode circuit, in series with R_5 , variation of the resistance

causes changes of potential in the same direction at the anodes of both valves. The variable resistor may, therefore, be used as a control to correct astigmatic distortion of the cathode ray tube. If this cathode inversion circuit is used to provide shift voltages and correction for astigmatic distortion, the potential of the final anode of the cathode ray tube should normally be about the same as the mean potential of the anodes of the two valves.

The cathode inversion amplifier may be used as a sum-and-difference device and as such finds frequent application in control and computing circuits.

Suppose that separate input voltages are applied to the grids of the two valves. Then the change of voltage developed between the two anodes is proportional to the difference between the input voltages, whilst the voltage developed at the cathode is proportional to the sum of the two input voltages. This assumes that the anode loads are equal and that the two valves have similar characteristics.

The cathode inversion circuit is often used to provide a paraphase output consisting of the result of adding two separate input pulses occurring at different times.

Fig. 382 shows the result of applying positive-going pulses at different instants to the grids of the two valves.

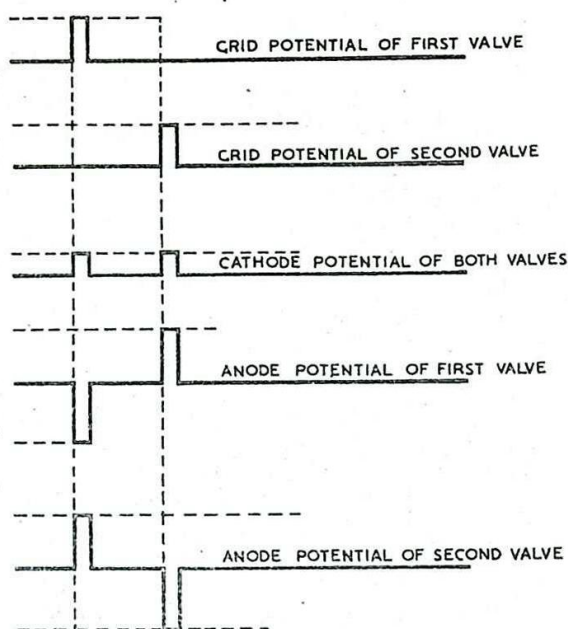


Fig. 382 - Use of a cathode-coupled paraphase circuit as an adding (subtracting) device.

LIMITATIONS OF THE USE OF VALVES AT HIGH FREQUENCIES

24. General

Amplifier and oscillator circuits employing negative-grid valves operate efficiently at frequencies up to a few hundreds of megacycles. As the frequency is increased efficient operation is more difficult to obtain. At frequencies of 1000 Mc/s or more the inherent difficulties are so great that amplification is not usually attempted. Further, at these frequencies negative-grid valves are inefficient in oscillator circuits and special valves (i.e. Magnetrons and Klystrons) are normally used as oscillators.

The difficulties experienced with amplifier and oscillator circuits employing negative-grid valves at the higher frequencies may be considered as due to :-

- (i) the finite time required for electrons to traverse the valve (transit time);
- (ii) limitations imposed by interelectrode and stray capacitance, inductance of valve leads and resistive and radiation losses.

25. Transit Time Effects

In RF Amplifiers the most serious consequence of a finite transit time is damping of the input circuit. At any frequency of operation there is at any instant, owing to the finite transit time, a difference in the number of electrons approaching and receding from the grid of the amplifier valve. The electrons forming the valve current at any instant induce a positive charge on the grid. If the grid voltage is increasing there is an excess of electrons approaching the grid, causing an increase of positive charge at the grid, i.e. there is a resultant current flowing into the grid from the external input circuit. If the grid voltage is decreasing there is an excess of electrons receding from the grid and this leads to a resultant current flow from the grid to the external circuit. Consequently an alternating current flows in the input circuit. At low frequencies of operation the transit time is small compared with the period of oscillation and the instant of zero grid current practically coincides with that of maximum grid voltage. The grid current leads the applied grid voltage by approximately 90° and there is practically no loss of power in the input circuit. There is however, a small increase in the grid-cathode capacitance in the valve when a signal is applied to the grid; this is unlikely to be of practical importance. At higher frequencies of operation the phase angle between the applied voltage and grid current is reduced, since there is an appreciable lag between the application of the grid voltage and the corresponding flow of current in the valve. Consequently a resistive component is introduced into the input impedance of the valve so that there is a loss of power in the input circuit. An approximate relation between the input resistance and transit time can be obtained as follows :-

let the grid-cathode voltage variation be given by :-

$$v_g = \hat{v}_g \sin \omega t.$$

If T_k = transit time of the cathode-grid space and T_a = transit time of the grid-anode space the current flowing towards the grid is given by :-

$$i_1 = G_m \hat{v}_g \sin \omega (t - T_k),$$

while the current flowing from the grid is given by

$$i_2 = G_m \hat{v}_g \sin \omega (t - T_k - T_a).$$

Consequently an alternating grid current flows given by :-

$$\begin{aligned} i_1 - i_2 &= 2 G_m \hat{v}_g \cos \omega \left(t - T_k - \frac{T_a}{2} \right) \sin \omega \frac{T_a}{2} \\ &= 2 G_m \hat{v}_g \sin \omega \frac{T_a}{2} \left(\cos \omega t \cos \omega \left(T_k + \frac{T_a}{2} \right) \right. \\ &\quad \left. + \sin \omega t \sin \omega \left(T_k + \frac{T_a}{2} \right) \right). \end{aligned}$$

The magnitude of the component of current in phase with v_g is given by :-

$$\hat{i}_r = 2 G_m \hat{v}_g \sin \frac{\omega T_a}{2} \sin \omega \left(T_k + \frac{T_a}{2} \right)$$

whilst that of the current in quadrature with v_g is given by :-

$$\hat{i}_c = 2 G_m \hat{v}_g \sin \frac{\omega T_a}{2} \cos \omega \left(T_k + \frac{T_a}{2} \right).$$

If the ratio of each transit time to the period of operation is small i.e. if ωT_k and ωT_a are small,

$$\hat{i}_r \doteq G_m \hat{v}_g \omega^2 \left(T_a T_k + \frac{T_a^2}{2} \right)$$

$$\text{and } \hat{i}_c \doteq G_m \hat{v}_g \omega \cdot T_a.$$

Hence the input resistance R_i is given by :-

$$R_i = \frac{1}{G_m \omega^2 \left(T_a T_k + \frac{T_a^2}{2} \right)}$$

and the increased input capacitance when the amplifier is operating is given by :-

$$C_i = G_m T_a.$$

The transit times are dependent on the linear dimensions of the valve and on the grid and anode voltages. The loss of power at the grid is used in increasing the mean electron velocity and appears as heat at the anode. The increase in input capacitance is small and can usually be neglected. However R_i , which damps the input tuned circuit, is inversely proportional to the square of the operating frequency, and hence the input conductance is proportional to the square of the frequency.

Consider the case of an amplifier circuit, the valve of which is a CV1091. The values of R_i at different frequencies are

100,000ohms at 10 Mc/s

4,000ohms at 50 Mc/s

250ohms at 200 Mc/s.

Thus if the dynamic resistance of the input tuned circuit is, say, 100k Ω the damping introduced is moderate at 10 Mc/s but prohibitive at 200 Mc/s unless special measures are taken. Hence, while damping of the input circuit is negligible at low frequencies it is considerable at frequencies above about 50 Mc/s.

Input damping can be reduced by decreasing the transit times; i.e. by reducing the spacing of the electrodes of the valve. The electrode areas must also be reduced so that the interelectrode capacitances of the valves are not increased. Extremely small spacing is the main criterion of Acorn valves, but the percentage of rejects in manufacture is high, and other valves (CV1091 and CV1065) with rather larger spacing are in general use. The grid-cathode transit time can be reduced by increasing the effective voltage at the grid. This means, however, that in order to maintain space-charge limitation, the emission density of the cathode must be high. Analysis shows that the cathode emission density required to ensure space charge limitation for a given ratio of transit time to operating period is proportional to the cube of the operating frequency.

Typical figures are as follows :-

If the transit time of the grid-cathode space must not be greater than a quarter of the operating period, then for a grid-cathode spacing of 1mm, the minimum current densities required at various frequencies are

$$2 \times 10^{-6} \text{ A/sq cm at } 10 \text{ Mc/s}$$

$$250 \times 10^{-6} \text{ A/sq cm at } 50 \text{ Mc/s}$$

$$0.016 \text{ A/sq cm at } 200 \text{ Mc/s}$$

$$2 \text{ A/sq cm at } 1000 \text{ Mc/s.}$$

In a valve oscillator the phase difference between grid voltage and anode current due to an appreciable transit time makes it difficult to adjust the relative phases at anode and grid for efficient operation. This does not apply to a valve amplifier, in which anode and grid circuits are independent of each other.

A further effect of transit time is an increase in anode dissipation. This may be of importance in power amplifier and oscillator circuits, in which the output power is limited by permissible anode dissipation. The Door-Knob valve is designed to give useful power output at frequencies as high as 600 Mc/s. The electrode spacing is small, and 35 watts can be dissipated at the anode since it is fitted with three radiation fins and enclosed in a relatively large glass envelope (of door-knob shape).

26. Circuit Limitations

(i) Frequency limitation

The values of the interelectrode capacitances and of the inductance of the internal leads of a valve operating as an oscillator fix the upper limit of frequency. If, however, the valve leads form part of a transmission line system of total length l , this frequency limitation is removed. For example, if the end of the transmission line is short-circuited, the resonant frequency of the circuit formed by the line and the interelectrode capacitance is given by

$$\tan \frac{2\pi l f}{c} = \frac{1}{2\pi f C_0 R_0} \quad \left(\begin{array}{l} \text{See Chap. 4 Sec. 14} \\ \text{for the effective} \\ \text{inductance } L_l \end{array} \right)$$

$$\text{where } c = 3 \times 10^{10} \text{ cms/sec}$$

C_0 = interelectrode capacitance

R_0 = characteristic resistance of the line.

For a given frequency f the permissible length of the line can be increased by decreasing R_0 . Hence a valve designed for use with coaxial lines (which are of low characteristic impedance), and whose electrodes and leads inside the envelope are integral parts of the lines, can be used for operation at very high frequencies.

In an oscillator the maintenance of continuous oscillations depends on the transference of energy from some source of power to a tuned circuit. In the case of a parallel tuned circuit fed with energy from a parallel negative-resistance source, the dynamic resistance of the tuned circuit must be greater than the magnitude of the negative resistance for oscillations to build up. The dynamic

resistance $\frac{L}{CR}$, where R is the series resistance of the coil L , may be made large only if C is kept sufficiently small. Hence it is necessary to keep interelectrode and stray capacitances to a minimum. This is still true when resonant lines are used as tuned circuits, since normally the line acts as an inductive susceptance resonating with interelectrode and other unavoidable capacitances.

The same problem does not arise in RF amplifiers above about 50 Mc/s since, owing to the high conductance of the input circuit due to transit time, cathode lead inductance, etc., the unavoidable resistance in parallel with the tuned circuits is much smaller than the undamped dynamic resistance of the latter. Since we may therefore neglect series damping owing to the low value of R_p , the parallel damping resistance, the Q of the total parallel circuit may be written

$$Q = R_p \sqrt{\frac{C}{L}}.$$

It may therefore be advantageous to make $\frac{C}{L}$ large in order to make the circuit sufficiently selective.

(ii) Input damping

The inductance of the internal cathode lead of a valve, which forms part of the input circuit of an amplifier, may bring about damping of the input circuit.

This inductive reactance in the cathode lead introduces feedback and modifies both input and output characteristics of the amplifier. In particular a component of the input current in phase with the applied voltage is generated, thus adding to the input conductance of the amplifier.

Fig. 383 shows the input circuit of an amplifier. An external voltage $v_i = \hat{v}_i \sin \omega t$ is applied to the grid, and a voltage v_L is developed across the cathode-lead inductance L_K . The cathode current is the sum of i_i and i_a , where i_i is the current through the grid-cathode capacitance C and i_a is the space current of the valve. Assuming the valve to be a pentode in which the slope resistance is very large compared with the anode load, we may write

$$i_a \approx G_m v_i.$$

It follows from the figure, that

$$v_s = v_i + v_L;$$

$$\text{hence } v_s = v_i + j\omega L(i_i + G_m v_i)$$

$$\text{or } v_s = \frac{-ji_i}{\omega C} + j\omega L(i_i + \frac{G_m ji_i}{\omega C});$$

i.e.,

$$v_s = \frac{i_i}{\omega C} \left[\omega L G_m + j(\omega^2 LC - 1) \right].$$

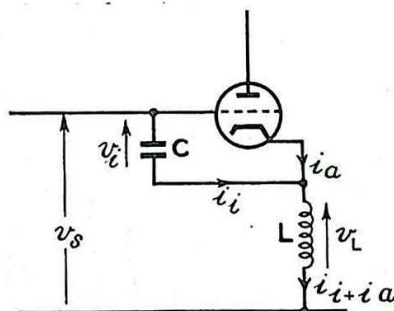


Fig. 383 - Effect of Cathode Lead Inductance.

Therefore the input admittance of the amplifier is given by

$$\frac{i_i}{v_s} = \frac{\omega C \sqrt{\omega L G_m - j(\omega^2 LC - 1)}}{(\omega L G_m)^2 + (\omega^2 LC - 1)^2}$$

Provided the operating frequency is not too high, then

$$\omega^2 LC \ll 1$$

$$\text{and } \omega L G_m \ll 1,$$

and the real part of the admittance (i.e., the conductance G_K) is given by

$$G_K \doteq G_m \omega^2 LC.$$

Thus the effect is that of placing in parallel with the input a resistance

$$R_K = \frac{1}{G_m \omega^2 LC}.$$

The input conductance $\frac{1}{R_K}$ is proportional to the mutual conductance and the square of the frequency just as the input conductance which results from transit time. The two effects are of the same order of magnitude, and the separation of the two causes is difficult.

Suppose, for example, that $G_m = 5$ mA/volt, $C = 5$ pF and $L = 0.05$ μ H, and the frequency of operation f is such that

$$\omega^2 LC \ll 1$$

$$\text{and } \omega L G_m \ll 1.$$

$$[\omega^2 LC \doteq 1 \text{ if } f \doteq 300 \text{ Mc/s and } \omega L G_m \doteq 1 \text{ if } f = 4000 \text{ Mc/s}]$$

Take $f = 50$ Mc/s.

$$\begin{aligned} \text{Then the value of } R_K &= \frac{1}{0.005 \cdot (2\pi \cdot 50 \cdot 10^6)^2 \cdot 0.05 \cdot 10^{-6} \cdot 5 \cdot 10^{-12}} \\ &\doteq 8000 \text{ ohms.} \end{aligned}$$

Since the pinch construction of supporting electrode leads in valves involves long leads, so that L is large, a decided decrease of input conductance is obtained if such a construction is replaced by short straight electrode leads mounted in a flat disc of glass. This latter type of base construction is used in the CV1091. A marked decrease of effective cathode lead inductance is obtained if such a lead consists of multiple wires in parallel as in the CV1136.

CHAPTER 8

VALVE OSCILLATORS

INTRODUCTION

1. Uses of Valve Oscillators

Oscillators are required in radar for use in many different types of circuits, to generate at frequencies which vary from about 4 c/s, in low-frequency switching circuits, to upwards of 10,000 Mc/s, for UHF equipments. Low frequencies are usually generated by relaxation oscillators, such as multivibrators or blocking oscillators. Radio frequency generators, from a few kc/s to 50 Mc/s, employ tuned circuits, usually of the "lumped" L-C type, with conventional valves. From 50-500 Mc/s specially constructed triodes and pentodes are used, the tuned circuits usually taking the form of lecher lines or tuned lengths of coaxial cable; otherwise the circuit designs are essentially the same as for lower-frequency generators. Above 500 Mc/s, novel valves which succeed because of transit-time effects, such as Magnetrons and Velocity-Modulated tubes of the Klystron type, replace valves of conventional design; tuned circuits, frequently in the form of cavity resonators, are wholly or partly built into such valves.

Although this chapter considers the action of valve oscillators mainly from the aspect of CW operation, the results quoted are generally still applicable when the oscillator is used as a transmitter in a pulse radar system. The times taken for the RF pulse to build up and to die away are usually small compared with the interval during which steady oscillations are produced, so that analyses based on CW operation are justified. The shape and duration of the oscillator pulse has, however, a very important bearing on receiver design. A harmonic analysis of the pulse reveals that a certain number of frequency components must be handled adequately by the oscillator output circuits if a good pulse shape is to be produced. The bandwidth involved is termed the Frequency Spread. For example, a 1 microsecond pulse would in practice have a frequency spread of at least 2 Mc/s. This would be diminished if the pulse duration were increased. (See Chap. 16 Sec. 2).

Although the modulating pulse has an important bearing on the frequency spread, other considerations, such as a poor rotating joint in the coupling circuit from transmitter to aerial, may also materially affect the output pulse shape. It is now general practice to examine the RF spectrum by means of a spectrum analyser, which is a super-heterodyne receiver whose local oscillator frequency is swept in synchronism with a CRO time base. The detected signals are fed on to the Y-plates to produce an amplitude-frequency spectrum of the components in the RF pulse.

2. Fundamental Requirements of a Valve Oscillator

Valve oscillators may generally be regarded as consisting of four component parts:-

- (i) A source of energy.
- (ii) A time-conscious circuit or frequency control device.
- (iii) A synchronous energy-feeding device.
- (iv) A load.

(i) is usually a DC power supply.

(ii) may be a resonant circuit or similar frequency-conscious device, such as a piezo-electric crystal or a phase-discriminating network. The time-conscious elements of a relaxation oscillator are of a different type, and these are dealt with in Chap. 10.

(iii) is normally a valve amplifier. The resonant circuit or other frequency-conscious device (ii) and the load (iv) dissipate energy, which must be provided for by an equivalent supply of energy from the power source (i).

Since (i) is a DC source and (ii) and (iv) constitute an AC load the power from (i) must be supplied at appropriate times, not only at the right frequency, but in the right phase with respect to successive cycles of oscillation. (This action corresponds closely to that of the escapement of a watch). The load, (iv) is an integral part of the oscillator and, in general, variations in its impedance will affect both frequency and amplitude of oscillations. Oscillators should therefore be designed so that the load is as little as possible affected by changes in the input impedance of any succeeding stage or other device. This can frequently be accomplished by inserting some form of buffer stage between the oscillator load and the output terminals. When this is done, changes in loading of the circuit as a whole affect the buffer stage only and do not disturb the loading of the oscillator to any appreciable extent. On the other hand, cases arise where the load is fed from the valve amplifier of the oscillator either directly or by coupling to the tuned circuit.

3. Conditions for Maintenance of Oscillations

The theory of oscillators is derived from that of their component parts, of which the resonant properties of tuned circuits and the amplifying properties of valves are of outstanding importance. In both cases the theory becomes complicated unless the alternating components of currents and voltages are assumed sinusoidal. For the derivation of initial conditions for self-maintenance of oscillations, these assumptions are valid; and, because of the relative simplicity of this method of analysis, useful criteria can be obtained which are serviceable in the design and adjustment of practical circuits.

For steady state conditions, however, this simple theory is unsatisfactory for the following reasons:-

- (i) All valve characteristics are non-linear, and the initial conditions for the maintenance of oscillations are valid only for infinitesimal oscillations. In all practical cases the amplitude of oscillation is finite and is determined, among other things, by the curvature of the valve characteristics.
- (ii) The curvature of the valve characteristics also necessarily results in distortion of the sinusoidal form of the current in the oscillatory circuit. In other words the production of harmonics is a necessary condition of self-maintained oscillations, and this affects the fundamental frequency of oscillation to some extent.
- (iii) Apart from changes in the effective resistance and reactance of the valve due to non-linearity of the characteristics, the interelectrode capacitances, which are included in the total reactances of the system, are not constant with respect to either the mean electrode potentials or the amplitude of the alternating voltages. .

These effects complicate accurate analysis, but overlooking them does not obscure the fundamental nature of the mechanism of oscillations.

In general, when the possibility of self-maintenance of oscillations is being considered in any circuit employing a negative-grid triode, a physical approach is often possible without detailed analysis, although this may be necessary as a last resort. To a first approximation the interelectrode capacitances C_{ga} , C_{gk} and C_{ak} of the valve may be considered as parts of the tuned circuit or circuits. The valve may then be replaced by a negative resistance, between anode and cathode. This is equivalent to assuming that anode and grid voltages are antiphase. Under these circumstances the frequency of oscillation does not depend on reactance effects of the valve current, and this is a desirable condition in most oscillators. When the valve is working in Class C, as in the case of power oscillators, the efficiency is liable to deteriorate considerably if the antiphase relation is substantially departed from. In any case, if anode and grid potentials are separated by a phase difference of less than 90° the valve acts as a positive resistance, and not as a generator of oscillations. This effect is illustrated in Fig. 384. Vectorially, assuming sinusoidal variation,

$$\vec{i}_a = G_m \vec{v}_g + \frac{\vec{v}_a}{R_a}$$

For the valve to act as a generator, ϕ , the angle between the current vector \vec{i}_a and the voltage vector \vec{v}_a , must be greater than 90° .

This is impossible either if $G_m \vec{v}_g < \frac{\vec{v}_a}{R_a}$, i.e., if $\mu \hat{v}_g < \hat{v}_a$, or if the phase difference between \vec{v}_a and \vec{v}_g is less than 90° .

Valve oscillators may frequently be regarded as positive feedback amplifiers, and such circuits may be employed in non-oscillatory conditions. A common example is the reaction amplifier of Fig. 385, which has essentially the same circuit as that of the tuned grid oscillator discussed in Sec. 4, but in which the coupling between anode and grid circuits is insufficient to cause self-maintenance of oscillations. The net effect is to increase the gain and selectivity (and the distortion) of the stage, without generating continuous oscillations.

Another way in which the regeneration of the feedback circuit may be regarded is illustrated in Fig. 386. A damped oscillatory circuit (a) rings when a square pulse is applied, as shown at (b). If further resistance is placed in parallel with the condenser the decrement is increased; (c). If the equivalent of a negative resistance (regeneration, or positive feedback) is applied to the output terminals, the decrement is reduced; (d). If sufficient positive feedback is provided, i.e., if the magnitude of the negative resistance is small enough, stable oscillations are maintained; (e). Whilst if more than sufficient feedback is present the oscillations increase in amplitude until limited by the inherent non-linearity of the circuit; (f).

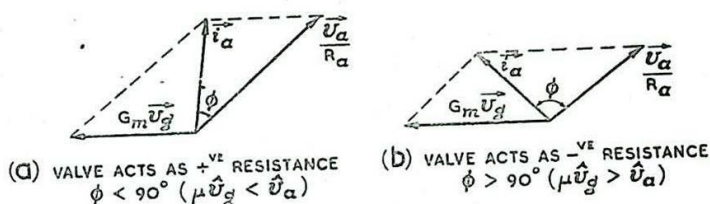


Fig. 384 - Regeneration criteria.

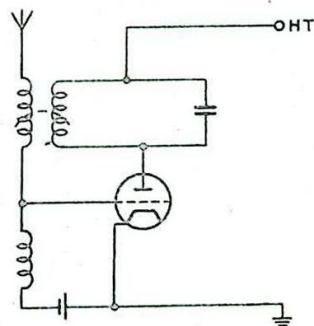


Fig. 385 - Reaction amplifier.

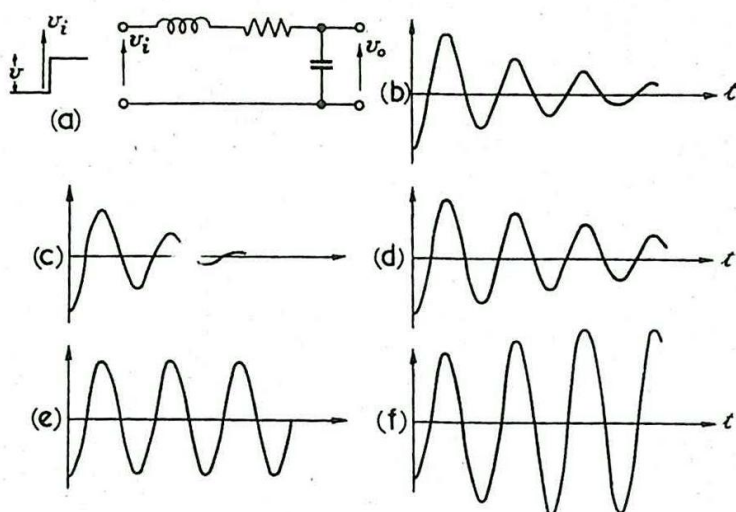


Fig. 386 - Effect of varying degrees of positive feedback on maintenance of oscillations.

OSCILLATORS EMPLOYING MUTUAL INDUCTIVE COUPLING

4. Tuned Grid Oscillator, Fig. 387

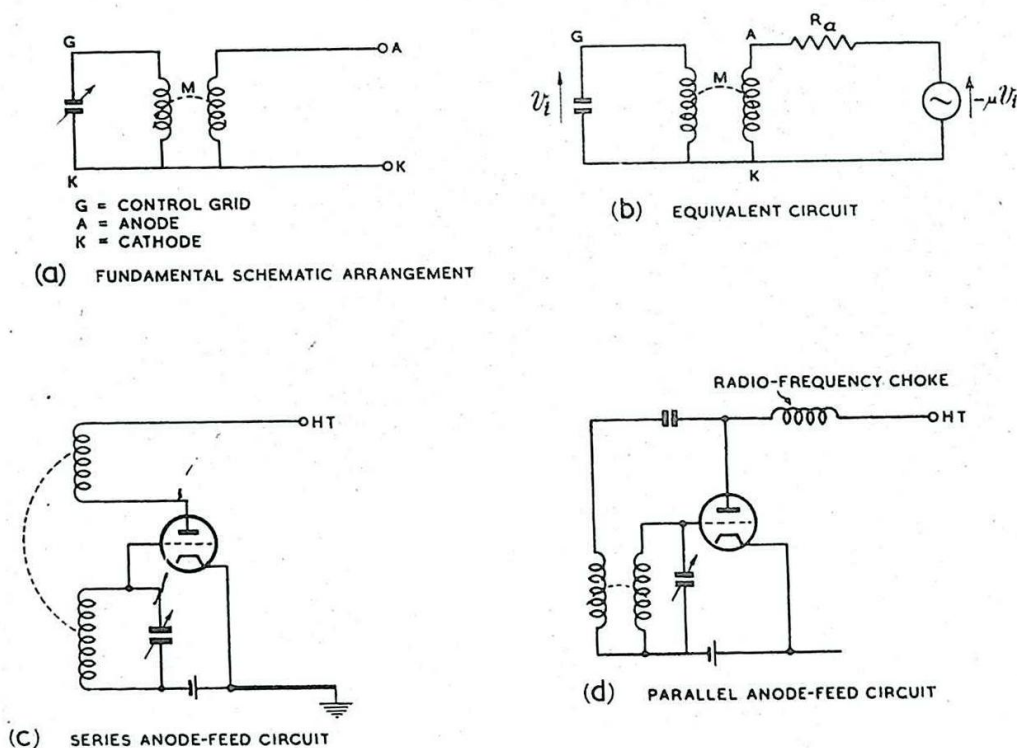


Fig. 387 - Tuned grid oscillator.

The fundamental arrangement is shown schematically at (a), and the equivalent circuit at (b). Feedback is provided by mutual inductive coupling M between anode and grid circuits. It may be shown that oscillations are maintainable at or near the resonant frequency of the tuned circuit. In this case the grid circuit is approximately resistive. If the anode circuit is also resistive, the condition for self-maintenance of oscillations is given by:-

$$M = \frac{1}{\omega_r Q G_d}; \text{ where } Q \text{ is the selectivity-factor of the}$$

tuned circuit, G_d the slope of the dynamic characteristic of the valve and $\omega_r = 2\pi f_r$, f_r being the resonant frequency of the tuned circuit, which is almost identical with f , the frequency of oscillations.

This relation, and those of a similar type applicable to other oscillators, are approximate only, and are based on sinusoidal oscillations of small amplitude under the simplest conditions. However, they provide an excellent guide to the modifications necessary to increase or decrease the likelihood of maintenance of oscillations.

Series and parallel anode-feeding arrangements are shown at (c) and (d).

5. Tuned Anode Oscillator, Fig. 388

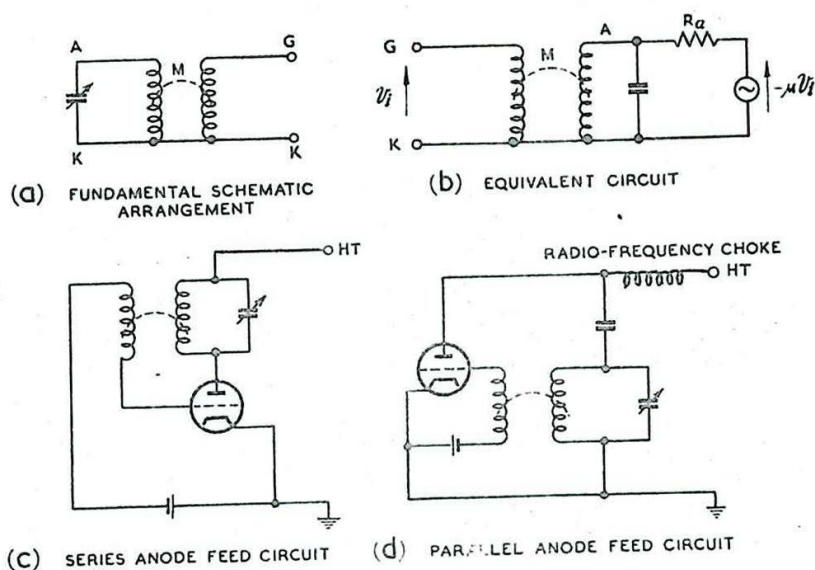


Fig. 388 - Tuned anode oscillator.

The fundamental arrangement is shown at (a) and the equivalent circuit at (b). The condition for self-maintenance of oscillations is the same as that for the tuned-grid circuit, and the modes of oscillation of the two circuits are very similar. Series and parallel anode-feeding arrangements are shown at (c) and (d).

6. Meissner Oscillator, Fig. 389

The fundamental arrangement is shown schematically at (a) and the equivalent circuit at (b). The condition for maintenance of oscillations is given approximately, in the simplest cases, by

$$\frac{M_a M_g}{L} = \frac{1}{\omega_r Q G_d}, \text{ where } M_a, M_g, \text{ are the mutual}$$

inductances between the coil L of the oscillatory circuit and the anode and grid coils respectively, there being no other coupling between the anode and grid coils. The other symbols are the same as for the tuned grid oscillator.

As in the analysis of the other circuits employing mutual inductive coupling, it is assumed that the sense of the coupling is such as to induce at the grid a voltage in phase with the anode current.

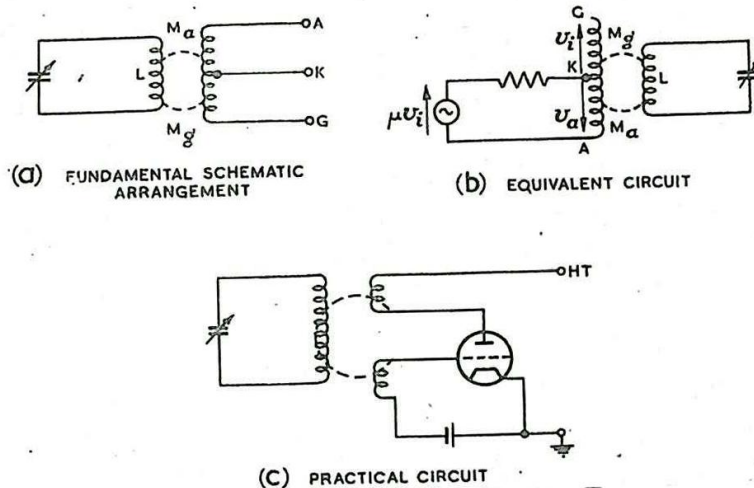


Fig. 389 - Meissner oscillator.

7. Magnetostriction Oscillator, Fig. 390

This circuit is closely analogous to the Meissner, the tuned circuit being replaced by a mechanically resonant rod of magnetic material inserted as a core in the anode and grid coils. As the field due to the current in the coils alternates such a rod varies in length at a frequency double that of the oscillator (assuming the rod is not polarised). This variation in length sets up mechanical oscillations which have a pronounced resonance, the resonant frequencies being of the order of 5-50 Kc/s in practice. The frequency stability is very good and the oscillator serves as a low frequency equivalent of the crystal controlled oscillator for use as a frequency standard in the supersonic range.

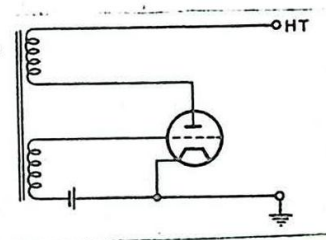


Fig. 390 - Magnetostriction oscillator.

OSCILLATORS EMPLOYING DIRECT COUPLING WITH A SINGLE TUNED CIRCUIT

8. Hartley Oscillator, Fig. 391

The fundamental arrangement is shown at (a) and the equivalent circuit at (b):

The requirement that the phase difference between anode and grid voltages should be greater than 90° is clearly met in this circuit provided high-Q components are used, and if resistances are neglected the ideal antiphase relationship is assured.

Under the simplest conditions, oscillations can be maintained provided $\frac{L_a L_g}{L_a + L_g} = \frac{1}{\omega_r^2 Q G_d}$, where G_d , ω_r and Q have their usual significance, and L_a and L_g are the inductances as indicated (the mutual coupling in this case is assumed to be negligible).

Because of the large impedance presented both at the grid and at the anode of the valve by the tapped-L arrangement at frequencies above resonance, this oscillator usually provides a high harmonic content.

Various series and parallel circuits are shown at (c) (d) and (e).

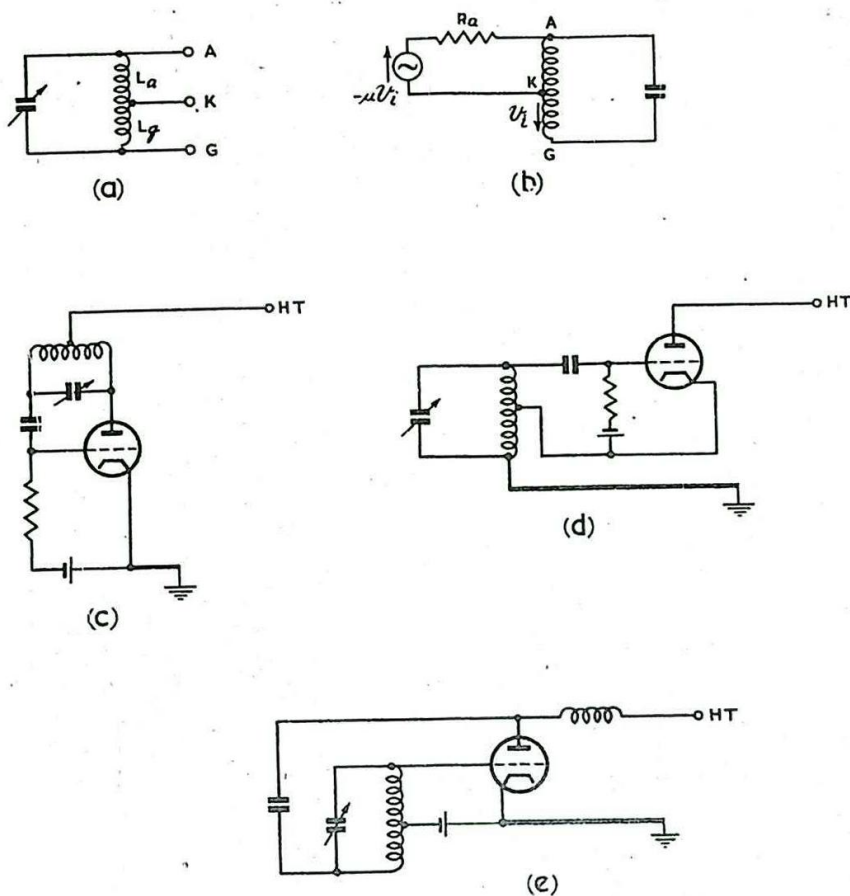


Fig. 391 - Hartley oscillator.

9. Colpitts oscillator, Fig. 392.

The fundamental arrangement is shown at (a) and the equivalent circuit at (b).

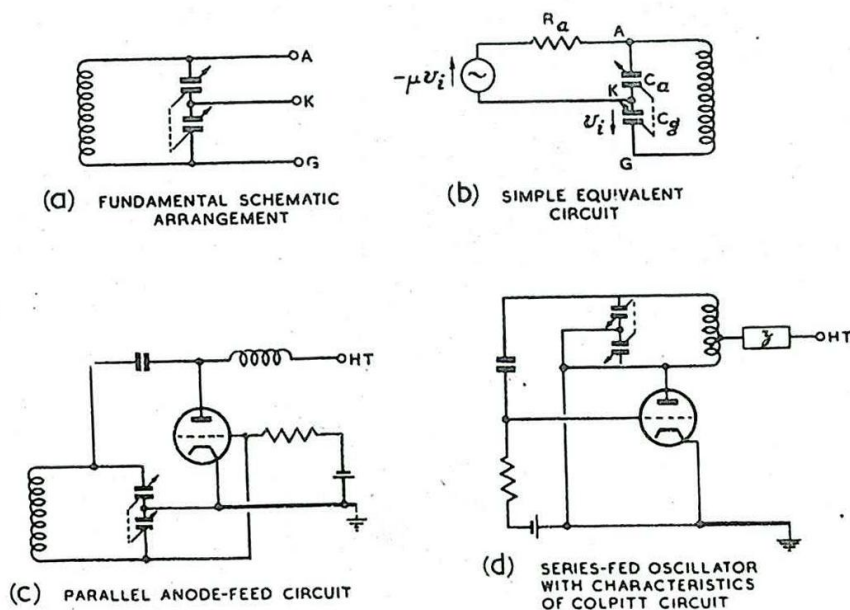


Fig. 392 - Colpitts oscillator.

This oscillator is analogous to the Hartley with the coils and condenser interchanged. It possesses good frequency stability, and is usable at higher frequencies than the Hartley circuit since stray capacitances, particularly C_{gk} and C_{ga} , are in parallel with the tuning capacitances, and thus do not upset the mode of oscillation.

Assuming the simplest circumstances, the condition for self-maintenance of oscillation may be written:-

$$G_d = \frac{\omega_r (C_a + C_g)}{Q}, \text{ where } G_d, \omega_r, Q \text{ have their}$$

usual significance and C_a and C_g are the capacitances as indicated in the figure.

The output is largely free from harmonics, since the tuned circuit presents a low impedance at both grid and anode for frequencies above the fundamental resonant frequency.

Practical circuits are shown at (c) and (d). There is no true series arrangement, but the circuit shown at (d) possesses the characteristics of a Colpitts oscillator, although the mode of oscillation is slightly more complex. The additional impedance z , usually an RF choke or small resistance, is necessary to prevent interference from the Hartley mode of oscillation. Without it oscillations may be impossible.

10. Dynatron Oscillator, Fig. 393

This is the simplest form of oscillator, requiring only two connections from the valve to the tuned circuit, (a). It relies for its mode of operation on the negatively-sloping portion of the $I_a - V_a$ characteristic of a tetrode, (b).

Over this region the valve acts as a negative resistance, and, provided the magnitude of this resistance is less than the dynamic resistance of the tuned circuit in parallel with it, oscillations are maintained.

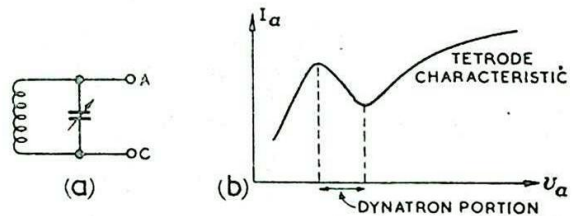


Fig. 393 - Dynatron oscillator.

11. Transitron Oscillator, Fig. 394

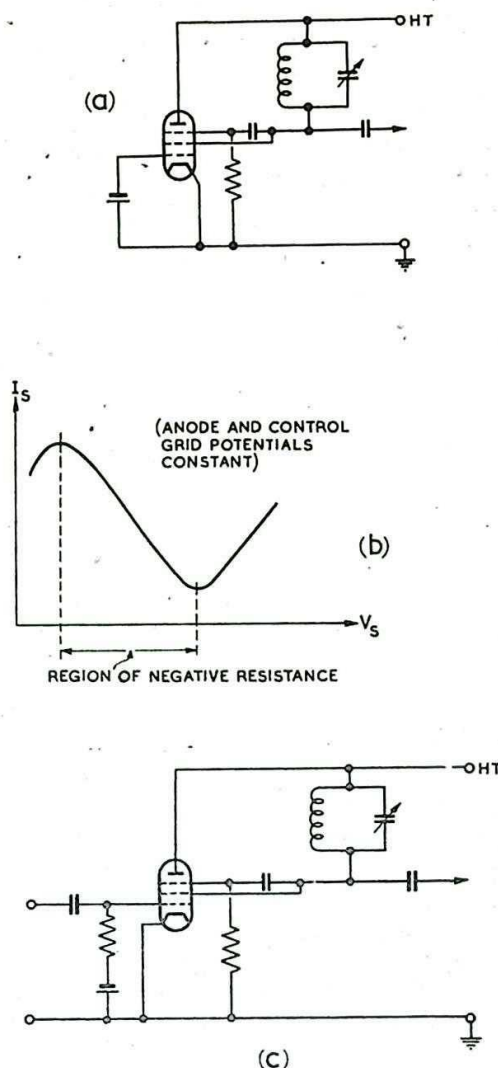
This type of circuit is more frequently employed as a relaxation oscillator or relay, and as such it is dealt with in Chap. 10. In essence, however, it has characteristics which are closely allied to those of the dynatron oscillator, and it may be used as a generator of sinusoidal oscillations.

The basis of operation of transitron circuits is discussed in Chap. 6 Sec. 34.

For certain values of anode and control grid potentials an increase in screen voltage gives rise to an increase in suppressor grid voltage which increases the anode current; this increase is drawn from the available space current, and, provided the mean potentials of the various electrodes are suitably chosen, is greater than the increase in space current due to the rise in screen voltage.

The net result is a decrease in screen current, so that the output resistance between screen and cathode is negative, (b). It may be convenient to use the control grid for triggering, as shown at (c). The valve oscillates while the input pulse holds the grid voltage above cut-off, the frequency being almost exactly that of the tuned circuit, assumed to incorporate the parallel stray capacitances present. After the grid potential falls below cut-off, oscillations die out, the time taken depending on the decrement of the circuit.

Fig. 394 - Transistron oscillator.



OSCILLATORS EMPLOYING DIRECT COUPLING WITH TWIN TUNED-CIRCUITS

12. General

The fundamental arrangement is shown schematically in Fig. 395(a). In such a circuit with loss-free components, oscillations can be generated under conditions which make the reactance of the coupling condenser C_c equal and opposite to the reactance of the two rejector circuits connected in series. For high-Q circuits any such oscillations would be damped out eventually, but would occur at substantially the same frequency as in the loss-free case.

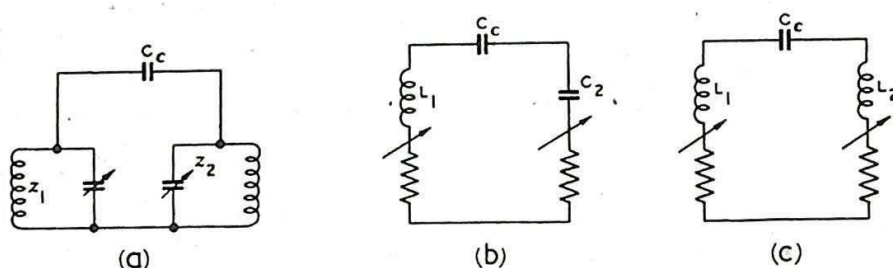


Fig. 395 - Compound oscillatory circuit.

At the oscillation frequency either one or both of the tuned circuits must be inductive, as shown at (b) and (c) respectively. The corresponding oscillation frequencies are then given sufficiently accurately by:-

$$(b) \quad f = \frac{1}{2\pi\sqrt{\frac{L_1 C_c C_2}{C_c + C_2}}} \quad \text{and} \quad (c) \quad f = \frac{1}{2\pi\sqrt{C_c (L_1 + L_2)}}$$

(The position of the arrows in Figs. 395(b) and (c) indicates that both of the series components of the impedance, the resistance and the reactance, are varied by adjustment of the corresponding tuned circuit of Fig. 395(a)).

By suitably connecting valve amplifiers to the networks of Fig. 395 the oscillations can be maintained. Figs. 396(a) and (b) show the types of vectorial relations one of which must apply to these circuits to permit a phase difference between anode and grid potentials relative to

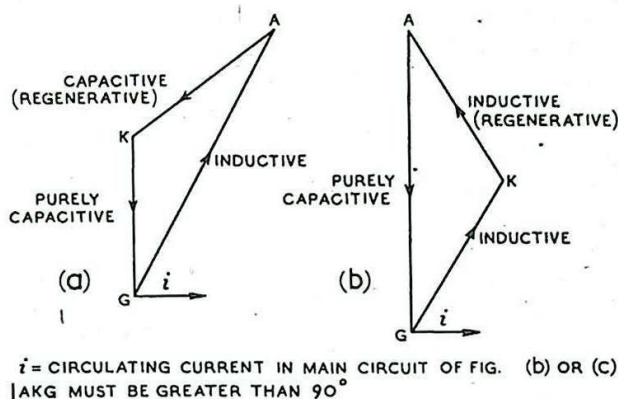


Fig. 396 - Twin circuit oscillator: vector diagram.

cathode of greater than 90° . The valve, acting as a generator between anode and cathode, makes the combined resistive component of the anode-cathode circuit negative; this is indicated in the vector diagrams by the word (Regenerative).

It follows that there are two arrangements of Fig. 395(a) which conform to Fig. 396(a) and one which conforms to Fig. 396(b); these are illustrated as equivalent circuits in Fig. 397(a), (b) and (c) respectively. They are classified as shown, according to the electrode which is common to the two tuned circuits.

The common cathode circuit is less useful than the others

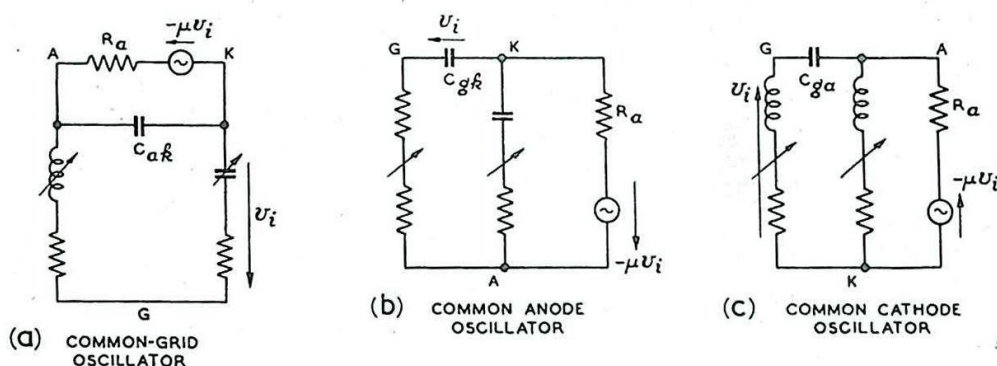


Fig. 397 - Twin circuit oscillator: equivalent circuits.

for very high frequency work because of the magnitude of the inter-electrode capacitance C_{ga} . This limits the oscillations to frequencies whose upper limit is of the order of a few hundred megacycles per second. Either of the other arrangements is capable of producing oscillations with normal triode valves up to 600 Mc/s, and with specially designed triodes up to 3000 Mc/s.

Common to all oscillators of this type is the advantage of having twin controls over the frequency and amplitude of oscillations. By suitable adjustment of the tuning controls both the amplitude and

frequency may be set to any desired values within the working range.

The mode of oscillation is similar to that of the Hartley oscillator in the circuit of Fig. 397(c) and to that of the Colpitts in circuits (a) and (b). In all cases, however, a much greater tuning range is available in the twin-circuit oscillator.

For a rejector circuit to be inductive it must be tuned to a frequency above the frequency of oscillation, and below if it is to be capacitive. Hence in the common cathode circuit the frequency of oscillation is below the resonant frequencies of both tuned circuits, whereas in the other cases it is intermediate between these two.

The behaviour of these oscillators may best be described in forms of modified circuits. Each tuned circuit is associated with certain components in parallel, the phase angle ϕ of the resultant circuit being a function of the selectivity and resonant frequency of the resultant circuit, and of the frequency of oscillation. It may be shown that, provided certain assumptions are justifiable, oscillations can be maintained at such a frequency that the resultant circuits are either both inductive or both capacitive, and the sum of these phase angles in each case is approximately 90° . Maximum amplitude occurs when both circuits are equally detuned, i.e. both phase angles are 45° . This occurs despite the fact that, as indicated in Figs. 397(a) and (b), one of the original tuned circuits may be inductive when the other is capacitive. This is because in these figures the addition of extra components for the purpose of simplifying the analysis has not been made.

13. Common-Cathode Oscillator

In this oscillator, the anode tuned circuit is associated with R_a , C_{ga} and C_{gk} in parallel, as shown in Fig. 398(a). The resonant frequency of this combined circuit is f_a and its phase angle

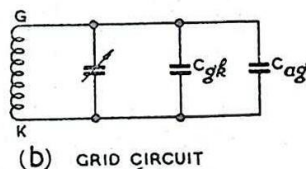
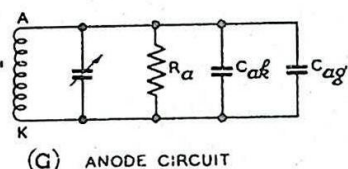


Fig. 398 - Common-cathode oscillator:
effective tuned circuits.

when the oscillation frequency is f is ϕ_a (inductive). The grid tuned circuit is associated with C_{gk} and C_{ga} in parallel, as shown at (b). The resonant frequency of the resultant circuit is f_g and its phase angle is ϕ_g (inductive). The results

$$(1) \phi_a + \phi_g = 90^\circ \quad \text{and}$$

$$(2) \text{ amplitude increases with } \sin 2\phi_a \text{ (which is equal to } \sin 2\phi_g)$$

depend upon the condition $G_m \gg \omega C_{ga}$.

14. Common-Anode Oscillator

In this arrangement, the grid tuned circuit is associated with C_{ga} and C_{gk} in parallel, whilst the cathode tuned circuit has the additional parallel components C_{ak} , C_{gk} , R_a and the resistance $\frac{1}{G_m}$ as shown in Fig. 399(a) and (b). It follows from the diagram of Fig. 397(b) that the addition of C_{gk} to the circuit between G and A changes

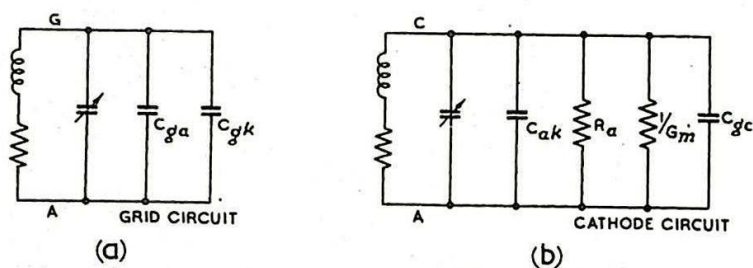


Fig. 399 - Common-anode oscillator:
effective tuned circuits.

this circuit from inductive to capacitive; this gives a good indication of the frequency of oscillation. The results

- (1) $\phi_g + \phi_k \approx 90^\circ$; (both ϕ_g and ϕ_k are capacitive) and
- (2) amplitude increases with $\sin 2\phi_g$; (which is equal to $\sin 2\phi_k$);

depend upon the condition $G_m \gg \omega C_{gk}$.

15. Common-Grid Oscillator

In this case the anode tuned circuit is considered in parallel with C_{ag} and C_{ak} , and the cathode tuned circuit has the additional

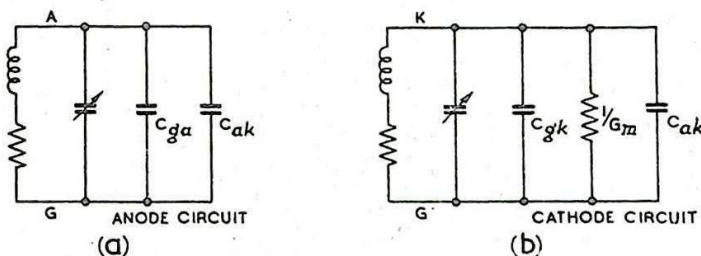


Fig. 400 - Common-grid oscillator:
effective tuned circuits.

parallel components C_{ak} , C_{gk} and the resistance $\frac{1}{G_m}$, as shown in Fig. 400(a) and (b). As in the previous case, the addition of C_{ak} to the circuit between anode and grid changes the impedance from inductive to capacitive, so that the frequency of oscillation lies between the resonant frequencies of the two circuits, with and without C_{ak} .

The results (1) $\phi_a + \phi_k \approx 90^\circ$ (both ϕ_a and ϕ_k are capacitive)
and

- (2) amplitude increases with $\sin 2\phi_a$
(which is equal to $\sin 2\phi_k$);

depend on the conditions $G_m \gg \omega C_{ak} \gg 1/R_a$.

16. Earthing the Twin-Circuit Oscillators

Owing to the earth capacitance of the various electrodes, these oscillators exhibit different properties for different earthing

arrangements. The anode is usually large and the stray capacitance to earth correspondingly high. Similarly, the necessity for the heaters to be close to the cathode tends to make the cathode-earth capacitance large.

In the common-cathode circuit it is usual to earth the cathode since this has least effect on the total capacitance between anode and grid. In the common-anode circuit either the cathode or anode may be earthed, but not the grid, as this effectively increases the coupling capacitance C_{gc} and also lowers the resonant frequency of the inductively tuned circuit. Both effects tend to lower the frequency of oscillation.

In the common grid oscillator, efficient oscillations may be obtained by earthing any of the electrodes. One arrangement, known as the Lighthouse Tube circuit, is of the common-grid, earthed-cathode type, and has given satisfactory performance at centimetre wavelengths. Valves with common-grid, earthed-grid circuits have also given good results. For high-power operation, where anode dissipation is considerable so that special precautions must be taken for cooling the anode, the common-grid, earthed-anode circuit is likely to be most useful.

17. Types of Tuned Circuits Employed

When a twin circuit oscillator is employed as a radar transmitting valve or as a local oscillator at frequencies of 200 Mc/s or more, the tuned circuits usually take the form of coaxial lines, and the valve is designed to fit onto a special coaxial assembly. Such an

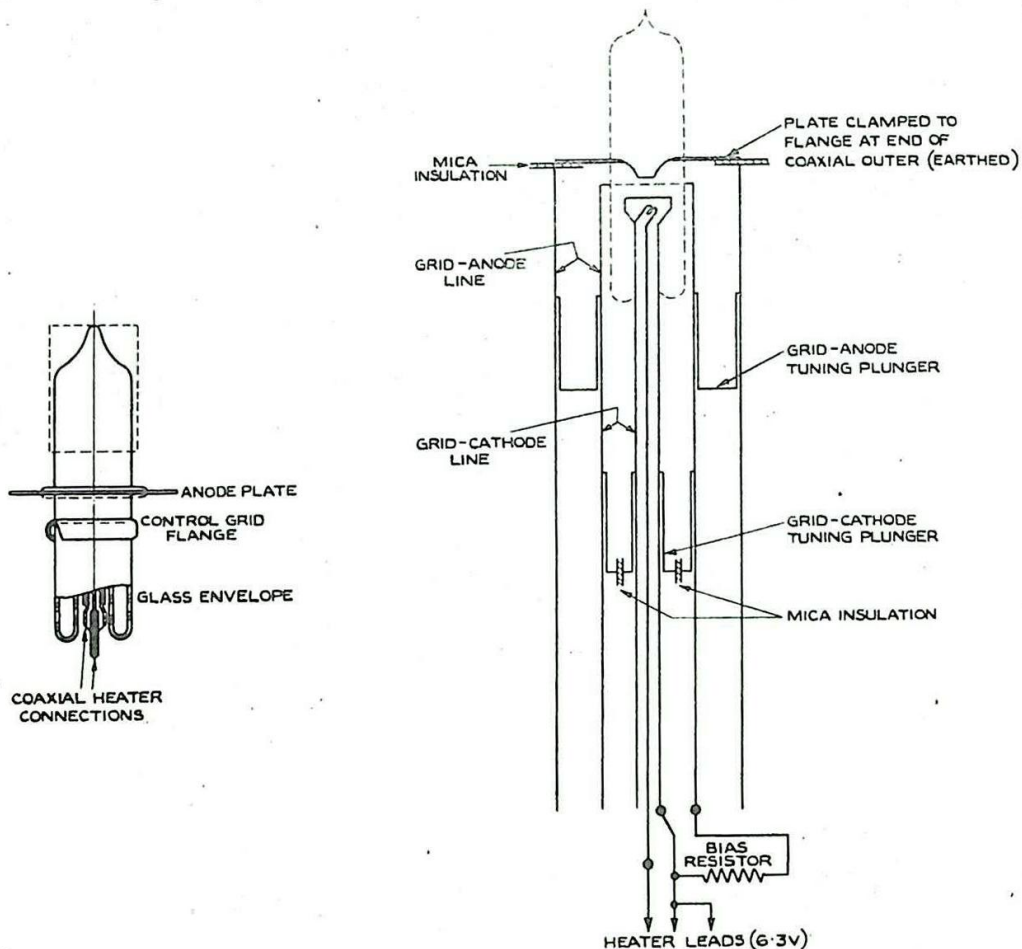


Fig. 401 - Negative-grid oscillator for centimetre wavelengths (CV 90).

arrangement is illustrated in Fig. 401. This shows a CV90 valve and the method of employing it as a common-grid oscillator. The anode is in the form of a flat plate sealed off on both faces from the glass envelope, and is suitable for clamping on to the end of a coaxial line system. The grid and heater connections are designed to fit various sections of coaxial tubing, the heater supply leads also being of coaxial construction. Anode-grid and grid-cathode lines are variable by means of plunger tuning. The anode is directly connected to the positive HT line, being insulated from the earthed outer of the coaxial line by a mica washer.

The valve, which measures 8.5 cm x 2.5 cm (neglecting the width of the anode plate) is capable of providing an RF output of 3 watts at 1000 Mc/s for an anode input of 10 watts. This output is reduced for higher frequencies, but by the use of harmonics the frequency can be raised to 5000 Mc/s, with an output of 0.1 watts. It may be used as a local oscillator or as a low-powered transmitter.

For lower frequencies conventional triodes with tuned open-wire lecher lines may be employed. (See Chap. 4 Fig. 147).

18. Crystal-Controlled Oscillators, Fig. 402.

Because of their excellent frequency stability, piezo-electric crystals are often employed in place of tuned circuits in oscillators of the twin-circuit type. The frequency is limited to a few megacycles, so that the magnitudes of the interelectrode capacitances are not of any great importance in determining which of the circuits should be used. Both the common-anode and the common-cathode arrangements shown in Figs. 402(a) and (b) are employed. In either

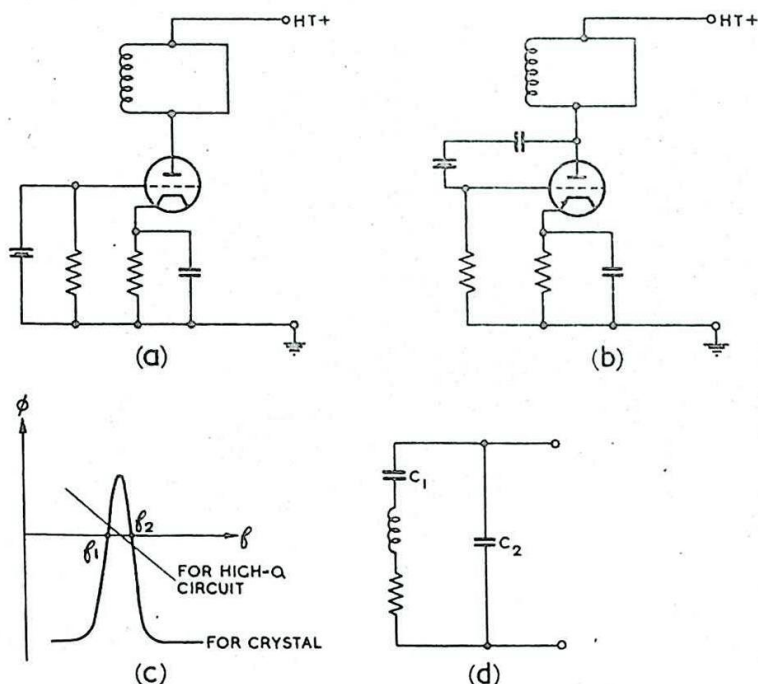


Fig. 402 - Crystal oscillator.

circuit the crystal must present an inductive impedance to the rest of the circuit for oscillations to be possible. Fig. 402(c) shows the variation of ϕ , the phase angle, for a typical piezo-electric crystal, and (d) shows the equivalent circuit. ϕ can be positive only between the frequencies f_1 and f_2 ; hence oscillations can occur only at a frequency greater than f_1 and less than f_2 . These limiting frequencies lie very close together, due to the fact that $C_2 \gg C_1$.

As indicated in the figure, ϕ alters very little over this range for the comparable case of a high-Q L-C circuit, which must therefore be considered less desirable in maintaining frequency stability.

Again, as $C_2 \gg C_1$, the effect of the external connections on the Q of the circuit, and on f_1 and f_2 is negligible.

For accuracies of the order of one part in a million or better the crystal should be kept at a uniform temperature by enclosure in a thermostatically controlled oven.

19. PUSH-PULL CIRCUITS, Fig. 403

The use of twin valve circuits symmetrically arranged makes the design of oscillators extremely simple. The anode and grid potentials of a conventional amplifier are normally antiphase; by connecting each anode to the grid of the other valve, the output of each valve provides the input for the other, in the correct phase for oscillations to be self-maintained. This is the essence of the multivibrator circuit described in Chap. 10. A similar circuit may be used to generate RF oscillations. The fundamental arrangement is shown schematically at (a) and a practical circuit at (b).

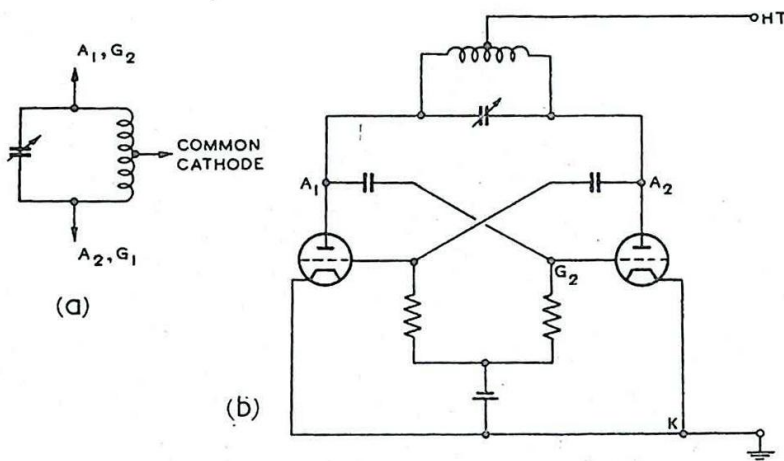


Fig. 403 - Push-pull circuit.

Push-pull circuits possess certain advantages over single-valve oscillators. Provided they are properly balanced, there are no even harmonics present in the output waveform. This conserves power and reduces interference with other transmitters. Since one electrode of each valve is connected to the corresponding electrode of the other, at least as far as radio frequencies are concerned, the interelectrode capacitances associated with this electrode and the tuned circuit are in series and the resultant effect on the tuned circuit is thereby halved. In the circuit shown at (b), for instance, the effective capacitance in parallel with the tuned circuit between the two grids is $\frac{C_{gk}}{2}$ where C_{gk} is the interelectrode capacitance between grid and cathode for either valve. This effect is nullified at very high frequencies, where the inductance of the cathode leads becomes of equal importance to the interelectrode capacitances. The doubling of this inductance in the push-pull circuit then destroys the advantage obtained by halving the capacitance.

20. RESISTANCE-CAPACITANCE OSCILLATORS, Fig. 404

A resistance-loaded, capacitance-coupled amplifier (a)

normally operates over the flat portion of its gain-frequency characteristic, where the phase difference between input and output is 180° . Over this region, neither gain nor phase is frequency-conscious. At low frequencies where the reactance of the coupling condenser is large enough, the gain falls off, and the phase of the output is advanced as illustrated in the vector diagram at (b). A similar falling-off of gain, accompanied by a phase lag, occurs at high frequencies due to the shunt reactance of the input capacitance of the succeeding stage and sundry strays. This is illustrated at (c).

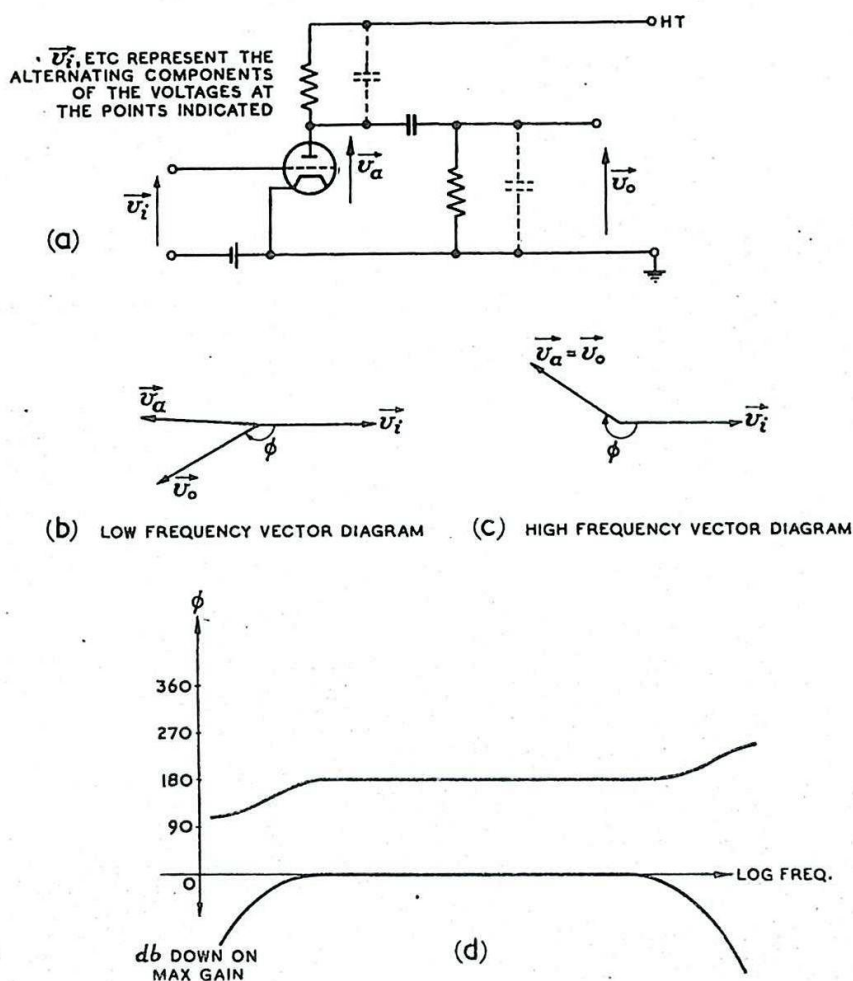


Fig. 404 - R-C oscillator.

The overall gain-frequency and phase-frequency characteristics of a typical R-C amplifier are shown in Fig. 404(d).

If two or more such amplifiers are used in cascade, the total phase shift is obtained by adding the individual phase shifts for the separate circuits. For example, if f_{120} is the frequency at which the input of each valve leads the output by 120° , the total phase-shift for three identical stages in cascade would be 360° . The output of the third stage would then be in phase with the input of the first, and if the output were coupled to the input a 3-valve oscillator of frequency f_{120} would be formed.

The corresponding two-valve circuit is not frequency discriminating over the linear portion of the valve characteristics, since the phase-shift of 180° per stage occurs over the flat portion of Fig. 404(d). Such a circuit usually works as a multivibrator, described in Chap. 10, and not as a generator of sinusoidal oscillations.

The number of valves required for a sinusoidal R-C oscillator may be reduced by the insertion of additional phase-shift meshes. Since the frequencies involved are usually low, it is normal to employ R-C rather than L-C networks, because of the prohibitive size of the latter when designed for use at these frequencies. A three-stage network such as that shown in Fig. 405(a) could be designed to operate with an amplifier providing 180° phase-shift. For oscillations to be maintained, the network would have to provide an additional 180° phase-shift, or 60° per section, whilst not introducing more attenuation than the available amplification.

The relative magnitudes of the components may be obtained from the formula of Chap. 3 Sec. 3, assuming the network properly terminated. Viz:- $\epsilon^{\gamma} + \epsilon^{-\gamma} = 2 + \frac{z_1}{z_2}$. On substitution we obtain

$$k (\cos 60^\circ - j \sin 60^\circ) + k^{-1} (\cos 60^\circ + j \sin 60^\circ) = 2 + \frac{1}{j\omega CR}$$

Hence $\frac{k}{2} + \frac{1}{2k} = 2$ and $k \frac{\sqrt{3}}{2} - \frac{1}{k} \frac{\sqrt{3}}{2} = \frac{1}{\omega CR}$,

whence we obtain $k = 2 + \sqrt{3}$ and $\frac{1}{\omega CR} = 3$.

From these results, $f = \frac{1}{2\pi \cdot 3CR} = \frac{1}{6\pi CR}$.

The attenuation due to the complete network is $k^3 \approx 52$. Hence the available amplification will need to be at least 52.

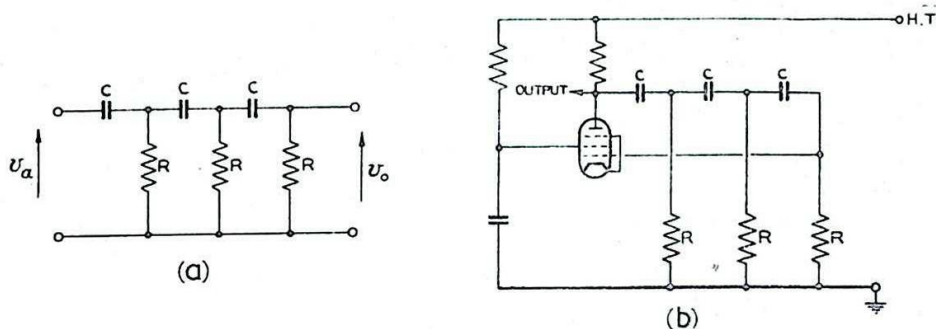


Fig. 405 - Single valve, three-mesh oscillator.

The three-mesh network may be replaced by a four-mesh circuit in which the phase-shift is 45° per mesh. This reduces the overall attenuation and enables a lower gain amplifier to be used. In practice the network is not normally properly terminated, so that the above calculation for f is not quite accurate. Either C or R of one of the meshes may be made variable for the oscillator to be tunable. A typical oscillator circuit is shown at (b).

Such a resistance-capacitance oscillator must operate in Class A for the output voltage to be reasonable sinusoidal. The bias is therefore usually provided by a cathode self-biasing network, since the disadvantages of this method of biasing do not arise under Class A working conditions.

An alternative form of R-C oscillator is shown in Fig. 406(a). The phase-discriminating feedback circuit takes the form of a Wien bridge, which is very sensitive to frequency changes. The vector diagrams of Figs. 406(b) - (e) illustrate the action. If the feedback link between Q and S is omitted, the voltages for the four arms of the bridge are as shown at (b). The insertion of this

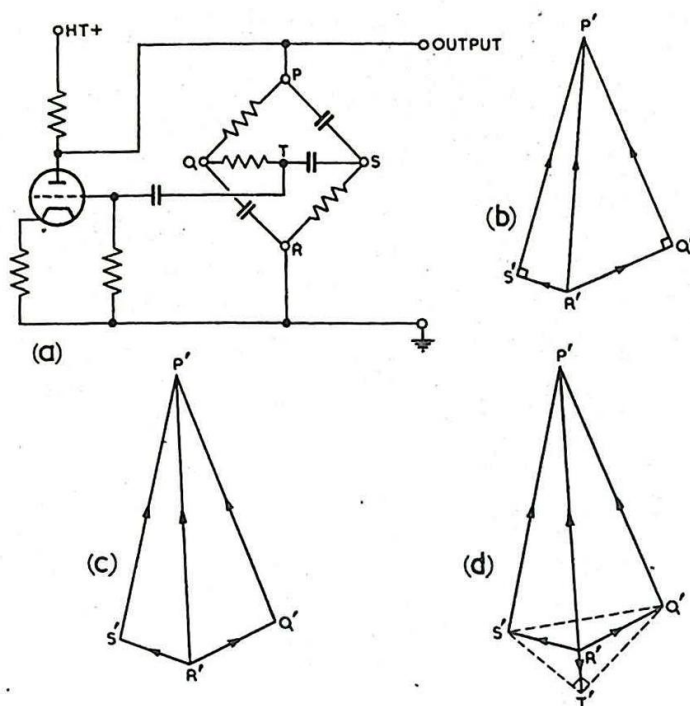


Fig. 406 - Wien bridge oscillator.

link distorts the right angles as shown at (c). The voltages developed across the components of the feedback link are given by (d), and the frequency of operation is such that $T'R'$ and $R'P'$ are collinear; i.e., that the anode and grid potentials are antiphase (with respect to cathode).

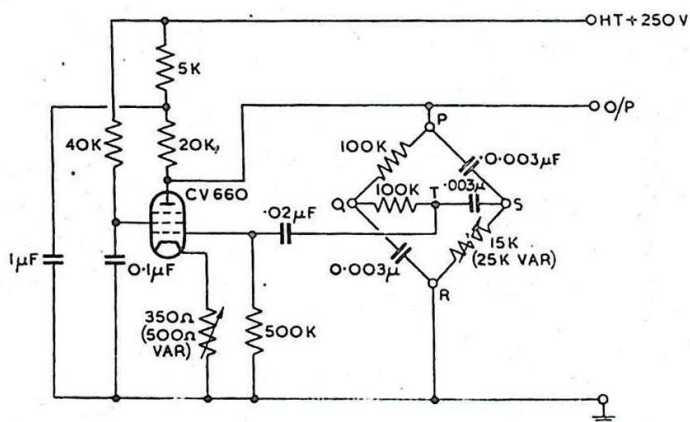


Fig. 407 - Wien bridge oscillator for use at 500 c/s.

Typical component values for an 800 c/s oscillator are shown in Fig. 407. Frequency is conveniently controlled by varying the small resistor (25 k Ω var.), whilst amplitude may be preset by varying the cathode feedback resistor (500 Ω var.).

VELOCITY MODULATED OSCILLATORS

21. General

Conventional valves are limited in their uses at UHF by the effects of transit time, reactance of electrodes and electrode leads, etc. (See Chap. 7 Secs. 25 and 26). While considerable success in minimising them has been achieved with specially constructed triodes,

having small electrodes very closely spaced, other types of valves which have utilised some of these effects have come into common usage. These depend for their mode of operation on variation of the velocities of the electrons (Velocity Modulation), instead of the total valve current (Density Modulation), as in conventional valves. Tubes employing this principle of velocity modulation must be hard, high-voltage tubes, so that the transit time for the passage of an electron through the modulating region is small enough to be comparable with the period of oscillations. The modulation depth is normally small, so that fluctuations at the high voltage end do not appreciably affect the initial velocity. This is analogous to the operation of a screened-grid valve, where variations in anode potential have little effect on grid modulation. Such a circuit is necessarily inefficient, since the tube current is largely DC, and the "anode" voltage is approximately constant, so that the power generated by the small AC components is at best only a few per cent of the total power dissipated. High efficiencies are obtainable only by modes of operation which are relatively complex, the description of which is beyond the scope of this work.

In practical velocity-modulated tubes the tuned circuits are normally of the resonant-cavity type, and are partly or wholly built into the valve envelopes. For simplicity they will be represented in the following sections as lumped circuits, and practical designs will be dealt with later.

22. Velocity Modulated Amplifier

Fig. 408 shows the basic velocity modulation circuit. An electron gun similar to that of a CRT projects a high velocity electron beam through an aperture in the plates of a condenser across which is developed an alternating field. The distance between the plates is assumed, in the first instance, to be small, so that the transit time between them is negligible; this will be considered in further detail later. When the resultant field is in one direction the beam is accelerated, and when in the opposite direction, retarded. Thus, after passing through the aperture of the first tuned circuit, or Buncher, the velocity of the beam is modulated, and the faster moving electrons tend to overtake the slower ones, so that Bunching occurs in the Drift Space; (Fig. 408).

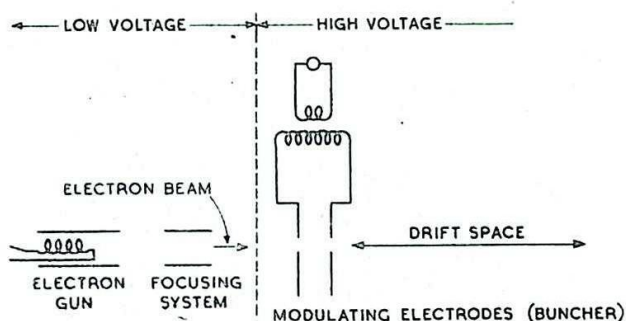


Fig. 408 - Velocity modulated circuit; schematic diagram.

The effect is illustrated in Fig. 409. For simplicity it is assumed that the modulation of velocity is sinusoidal, the field of the buncher imparting to the steady velocity u_0 a deviation velocity $u_d = \hat{u}_d \sin 2\pi ft$. $100 \frac{\hat{u}_d}{u_0}$ is the percentage modulation depth.

The resultant velocity $u = u_0 + u_d$ will be referred to as the Drift-Space Entry Velocity, or simply the Entry Velocity. The instant at which a "snapshot" of the electron beam distribution is taken is denoted by t , and t_0 is the instant at which the electron under consideration entered the drift space. At (a), u is plotted against $t - t_0$ for the case where $\sin 2\pi ft = 0$; i.e., $t = 0 + nT$ where $T = \frac{1}{f}$. At (b) the distance s which the electrons have travelled into the drift space is plotted against $t - t_0$. Thus (a) gives the entry velocity and (b) the snapshot distribution at the instant t , the

interval $t - t_0$ being the time spent in the drift space by the electron considered.

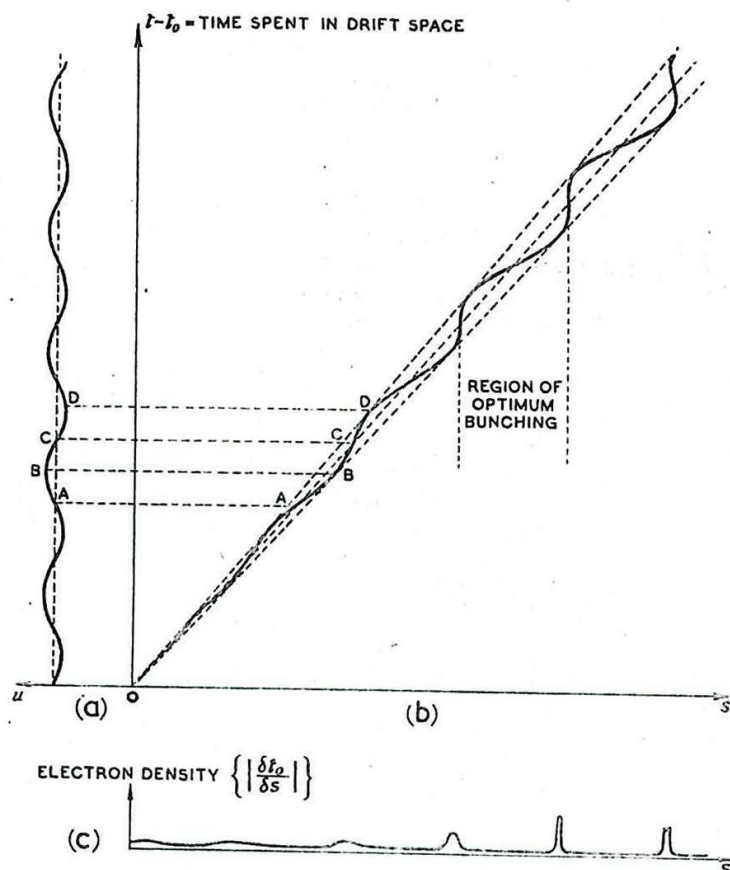


Fig. 409 - Drift space entry velocity and charge distribution ($t = \pi T + 0$).

The entry velocity of electrons emerging at instants A and C of Fig. 409(a) is u_0 , corresponding to the line OAC [$s = u(t - t_0)$] in Fig. 409(b). At B the entry velocity has a maximum value, $u_0 + u_d$, corresponding to the line OB [$s = (u_0 + u_d)(t - t_0)$], whilst at D the entry velocity is a minimum, corresponding to OD [$s = (u_0 - u_d)(t - t_0)$]. The position of an electron entering the drift space at any instant t_0 for which $\sin 2\pi f(t)$ has any value other than 0 or ± 1 may be found by sinusoidal interpolation between the three constant-velocity lines OB, OA, OD. It is assumed that the electron distribution along the axis $t - t_0$ is uniform, i.e. that equal numbers of electrons are emitted in equal intervals of time. The dispositions of these electrons are given by graph (b) for $\sin 2\pi f t = 0$, and the ratio $\left| \frac{\delta t_0}{\delta s} \right|$ (Fig. 409(c)) gives the mean electron density for the region of length δs . In the figure, δt_0 is taken as one twelfth of the period T .

Curves similar to those of Fig. 409(b) may be plotted for the instants $t = \frac{T}{4}, \frac{T}{2}, \frac{3T}{4}$, and these are shown in Fig. 410(a). The corresponding line distributions are at (b), (c) and (d). These figures show how the bunches are formed by the faster-moving electrons overtaking the slower ones, the bunches moving at approximately the mean beam velocity u_0 , so that they are separated by a distance $s = u_0 T$.

The regions in Fig. 410(a) where $\frac{\partial s}{\partial t_0} = 0$ are sometimes called Bunching Planes. These corresponding to maximum peaking effect in the line distribution diagrams (b), (c) and (d). It should

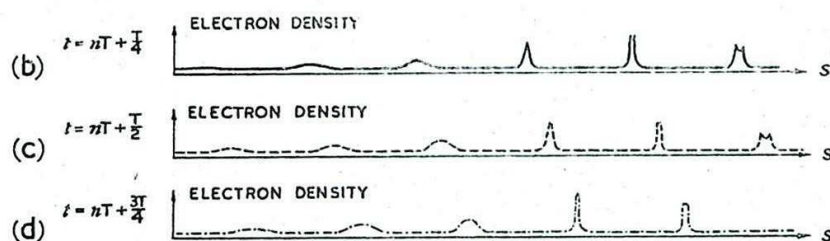
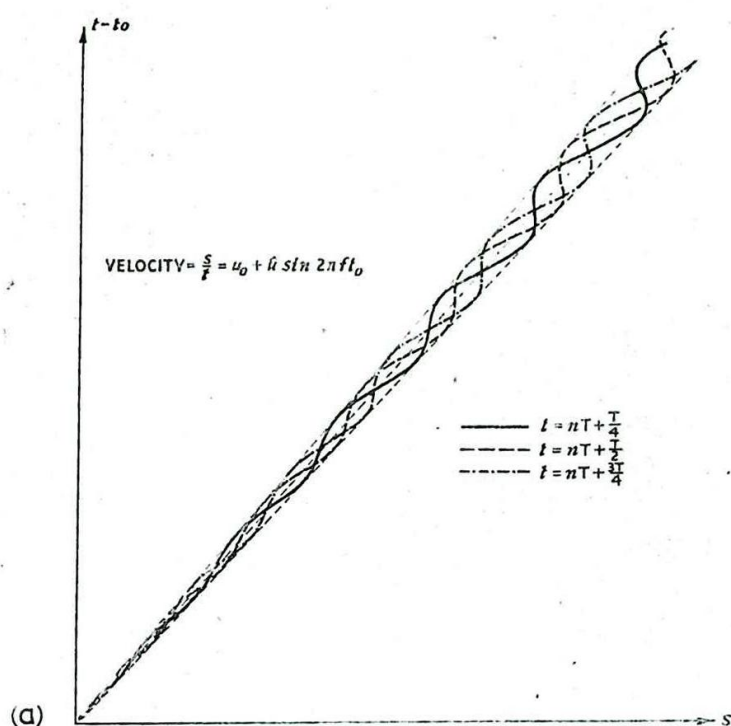


Fig. 410 - Drift space charge distribution for different values of t .

be obvious, however, that bunching is not confined to these particular regions, but occurs in varying degrees at all points of the drift space.

The effects illustrated are modified in practice by the following considerations which have been ignored in the simple diagrammatic treatment.

(i) The entry velocity does not vary sinusoidally with time. This does not affect the principle of the method, as the sinusoidal form of Fig. 409(a) may be replaced by any other, and the process applied in the same way.

(ii) Acceleration is not instantaneous, occurring at a point, but takes place over a period of time comparable with T ; one effect of this is that there may be some density modulation present which cannot be accounted for by velocity modulation only.

(iii) There is mutual repulsion between the electrons of the beam. This is usually a second order effect.

The manner in which modulation occurs merits further detailed

consideration, as it is fundamental in many centimetre-wave generators. An idealised modulating element is shown in Fig. 411 illustrating the instantaneous distribution of the electric field; the corresponding distribution of axial potential is shown in Fig. 412(a). It has been assumed that the region outside the plates is devoid of alternating fields; this is signified by the 0 - 0 lines within which these fields must be confined. Accordingly, whatever rise in potential occurs along the axis at any instant must be offset by a corresponding fall before the outside is reached. From energy considerations, it follows that an electron which instantaneously passes axially through this field meets with no net acceleration or retardation.

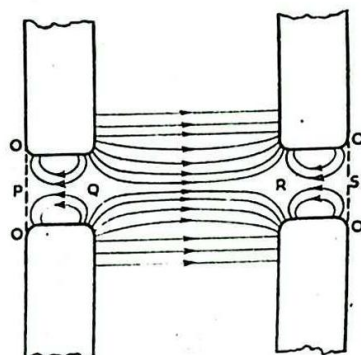


Fig. 411 - Instantaneous field distribution within the buncher gap.

Suppose, on the other hand, that the time of transit of the electron is comparable with the periodic time of the modulating voltage. Some electrons enter the region with the potential distribution as at (a), being accelerated from P to Q. By the time they pass Q, the direction of the field may have changed to that shown at (b), so that the electrons are further accelerated for the journey Q - R. Provided the distance QR is approximately $\frac{u_0}{2f}$, the field will again reverse when the electrons are in the region of R, so that the final stage of the motion, from R to S, is again accelerating.

Electrons which are a quarter-period ahead of these reach Q when the potential there is a maximum, and encounter a retarding field as they approach the centre. At this point the field everywhere is zero, and the electron velocity is again u_0 . The second half of the transit is a repetition of the first half, the electrons being accelerated until they reach R, at which point the potential is a maximum, and retarded from R to S, being ejected with velocity u_0 .

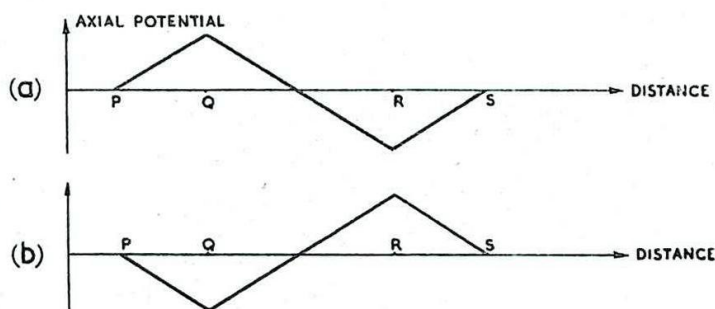


Fig. 412 - Instantaneous potential within the rhumbatron gap.

Electrons which are a half-cycle ahead of the first are correspondingly retarded throughout the motion. The behaviour of other electrons may be estimated by interpolation.

An accelerated electron is supplied with energy from the source which feeds the buncher. A retarded electron supplies this source with energy; so that the net energy lost by the buncher in this manner is, to a first approximation only, zero.

If the electrons after modulation are collected at an anode, or Catcher, to which is attached a parallel tuned circuit, as shown in Fig. 413, oscillations are set up in the latter in the same way as they are in a Class C amplifier with tuned circuit load.

At the anode it is the density modulation, due to the bunching, which is important, and not the difference in velocity itself.

If the current variation of Fig. 414 is compared with that of a Class C amplifier, the closeness of the analogy will be apparent.

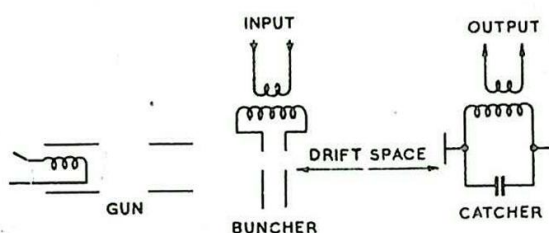


Fig. 413 - Extracting energy from velocity modulated circuit.

The mechanism thus described in schematic form constitutes an amplifier of sorts, since, to a first approximation, the input supplies no net power, whereas useful power is available at the output. Unfortunately, the amplification

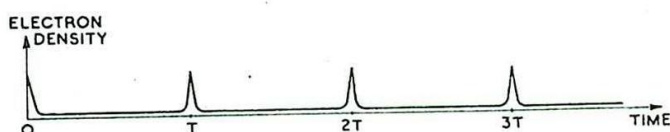


Fig. 414 - Variation of electron density with time in the region of optimum bunching.

obtainable from such a system is not independent of input amplitude, i.e. of the modulation depth. If the depth of modulation is doubled the output amplitude is not correspondingly doubled, and may be diminished. In fact the variation of amplification with modulation depth is extremely complex, depending on the length of the drift space, mean velocity and frequency. Although amplifiers employing the principles described are in use, they are very limited in their application and are inherently noisy. These disadvantages are not important when the methods are applied to the design of a constant frequency oscillator, and it is in this respect that they are commonly employed.

It is not essential for the bunched electrons to impinge on an electrode connected directly to a tuned circuit in order that energy may be transferred to the latter from the valve current. The modulated beam may be directed through the aperture of a second pair of electrodes similar to the modulating set, connected to a circuit

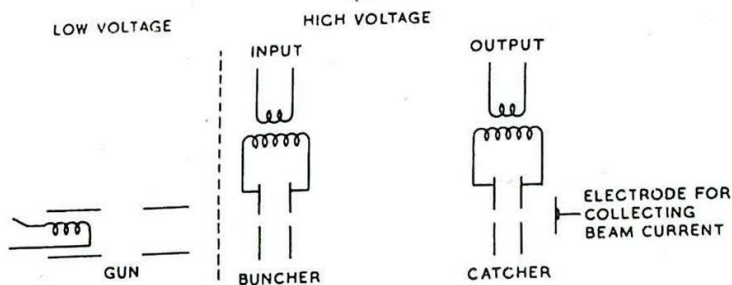


Fig. 415 - Alternative method of extracting energy.

tuned to the operating frequency (Fig. 415). Each bunch of electrons passing through the aperture will induce currents to flow in the circuit, and successive bunches will, because of the resonance of the circuit, arrive at the correct instants to reinforce oscillations already present there. In other words, the electron beam may be made to augment the alternating displacement current between the condenser plates without actually impinging on either of them. In practice it is probable that both effects occur.

To convert the double-tuned circuit amplifier into an oscillator all that is required is a feedback circuit from catcher to the buncher. This is indicated in Fig. 416. The phase of the feedback must correspond to the length of the drift space and the mean velocity of the beam for regeneration of positive feedback to occur at the desired frequency. This means that the feedback

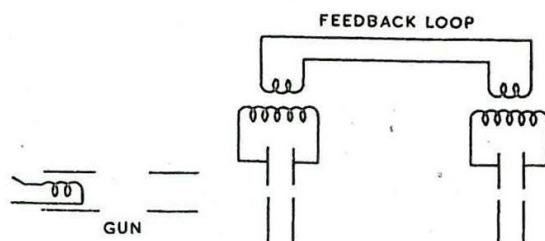


Fig. 416 - Adaptation of velocity modulated amplifier as oscillator.

length, since a difference in length of $\frac{\lambda}{2}$ would change this feedback from positive to negative or vice versa.

The foregoing description has been limited to theoretical circuits without considering the practical forms which these circuits may take. Common forms of practical velocity-modulated tubes will be described in the following paragraphs.

23. Velocity Modulated Oscillators in Common Use

Heil Tube, Fig. 417

This oscillator is an inefficient, low-powered generator possessing moderately good frequency stability. Its main advantage is that it may be tuned over a wide frequency range. The combined tuned-circuit and feedback link are formed of a short length of coaxial line with a hollow inner conductor. Apertures are cut at opposite ends of a diameter, the first to permit the entry of the electron beam, thus forming the buncher; the second allows the modulated beam to emerge, and constitutes the catcher, currents being induced in the inside walls of the coaxial line by the pulses of current in the beam. The drift space consists of the region in the neighbourhood of the diameter joining the two apertures. The transit time across this space must be such as to ensure that the currents induced at the output are in phase with the currents at the input. The emerging electrons are collected at an anode beyond the second aperture.

The resonant frequency of the circuit depends on the effective length of the coaxial line. This line is usually terminated in a variable length of short-circuited lecher line. As this length is varied, the voltage difference between cathode and accelerating electrode must be changed correspondingly to maintain the phase equality at input and output.

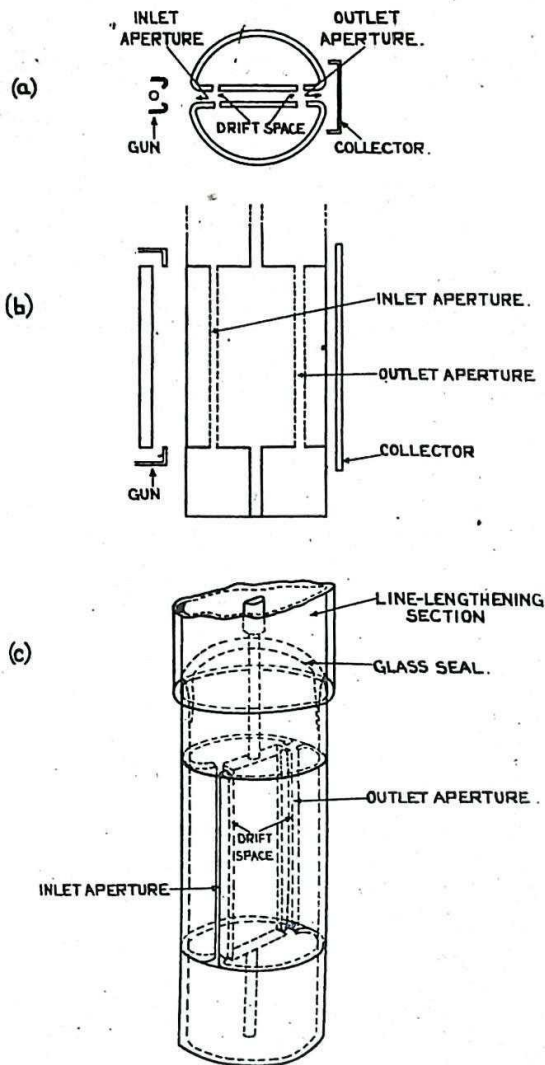


Fig. 417 - Heil tube

Double Rhumbatron Klystron, Fig. 418

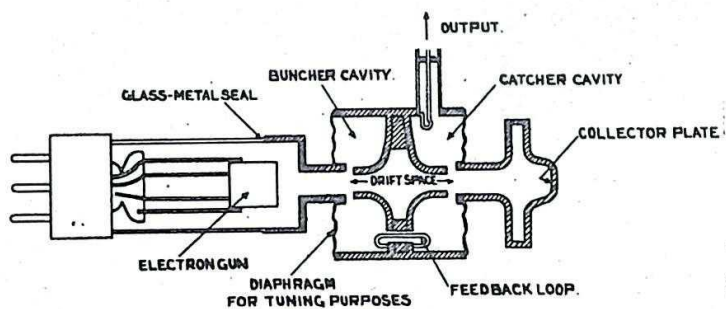


Fig. 418 - Double rhumbatron klystron.

In this tube the two tuned circuits are of the resonant cavity type, the rhumbatron shape being particularly suited to the aperture method of feeding and extracting energy. The term Klystron is usually applied to a velocity modulated oscillator using a

cavity resonator. Feedback is provided by a coaxial cable, terminated at both ends by loops which are inductively coupled to the cavities. The length of this cable must correspond to the length of the drift space between the apertures of the two rhumbatrons. Owing to the finite size of the loops and the variation of their input impedance with frequency, the optimum cable length will not be independent of frequency.

A major difficulty encountered in this oscillator is the necessity for the two circuits to have the same resonant frequency. The high selectivity of the rhumbatrons used permits only the smallest deviation from resonance before the maintenance of oscillations becomes impossible.

Sutton Tube, or Reflector Type Klystron

In the Sutton Tube, or reflector type of klystron, instead of a second tuned circuit and aperture being employed, the electron beam is reflected and made to return to the modulating aperture at the correct phase for the maintenance of oscillations.

This is done by the insertion of a reflecting plate at a potential negative to cathode (Fig. 419). Because of the considerable retardation and subsequent concentration of electrons between rhumbatron and reflector, the inaccuracies of neglecting interaction between electrons are greatly increased; but the simplified explanation which follows, although it neglects this, is sufficient to indicate the general mode of operation.

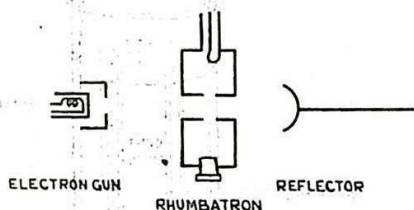


Fig. 419 - Schematic diagram of Sutton tube.

It is assumed as a first approximation that an electron after emerging from the aperture is subject to a constant retarding force E . This force brings the electron to rest and causes it to return to the aperture with the same velocity as that with which it emerged. The method by which bunching occurs under these conditions is illustrated in Fig. 420. The notation used in this diagram is the same as that of Fig. 409. The velocity u of the returning electrons varies throughout the drift space as shown at (a).

Because of the constant retarding force, the straight lines OB, OA, OC, of Fig. 409(b) become the parabolas shown in Fig. 420(b). The line distribution of electrons for different values of t is given by the four curves for $t = 0, \frac{T}{4}, \frac{T}{2}, \frac{3T}{4}$. The diagram shows

that at the instant $t = 0$ the electron density $\frac{\delta t_0}{\delta s}$, at the aperture (where $s = 0$) due to the reflected beam, is large, whereas at all other instants $\frac{\delta t_0}{\delta s}$ is small. Approximately half the electrons pass through the aperture in a "bunch" whilst the passage of the remainder is spread over the rest of each cycle.

If e is the charge, and m the mass of an electron, the retardation r due to the field E is $r = \frac{Ee}{m}$. The time for an electron to be brought to rest and to return again to the aperture

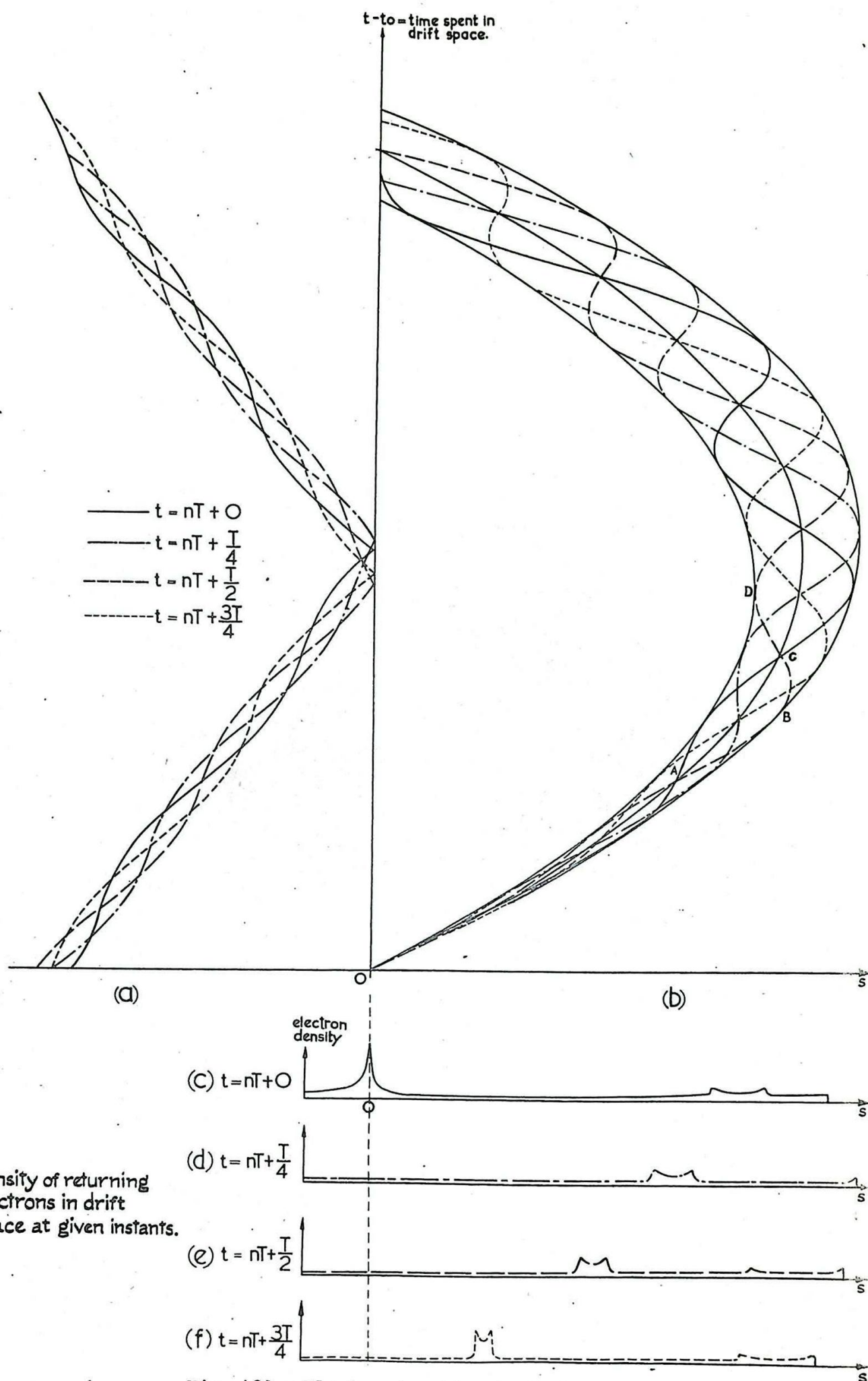


Fig. 420 - Electron bunching in a reflector klystron.

is $t = \frac{2u}{r}$. The mean time is therefore $\bar{t} = \frac{2u_0}{r} = \frac{2mu_0}{Ee}$.

For oscillations to be maintained there must be a suitable phase relationship between the return of the bunches to the aperture and the oscillatory field which causes the bunching; i.e., \bar{t} must have a suitable value, dependent on the manner in which energy is fed into the cavity.

Further, for optimum bunching the difference between the times spent in the drift space by the fastest and the slowest electrons should be approximately an odd multiple of $\frac{T}{2}$. The time taken to bring to rest an electron which enters the drift space with velocity u is $\frac{u}{r}$; the total time it spends in the drift space is therefore $\frac{2u}{r}$. Hence the time-difference for the fastest and slowest electrons is

$$\begin{aligned} 2 \frac{\hat{u}}{r} - 2 \frac{\check{u}}{r} &= 2 \frac{u_0 + \hat{u}_d}{r} - 2 \frac{u_0 - \hat{u}_d}{r} \\ &= 4 \frac{\hat{u}_d}{r} \end{aligned}$$

Hence $4 \frac{\hat{u}_d}{r} = (p + 1) \frac{T}{2}$ where p is an integer

$$= \frac{(p + 1)}{2f}$$

$$\text{or } \hat{u}_d = \frac{p + 1}{8f} \cdot \frac{Ee}{m}$$

In order to set up the oscillator at a given frequency, so as to produce the maximum amplitude it is necessary to adjust both \bar{t} and \hat{u}_d ; i.e., both u_0 and E . Thus both resonator-cathode and resonator-reflector voltages must be variable.

Sutton tubes are still undergoing development, and details of design are not standardised. A type of tube in common use is shown in Fig. 421.

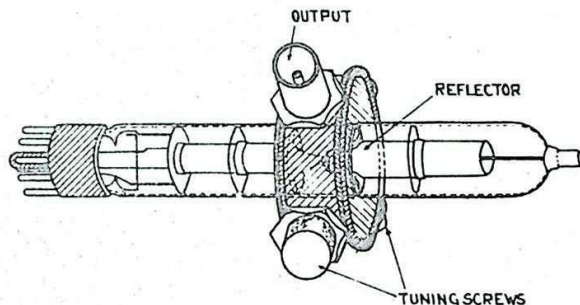


Fig. 421 - Sutton tube (CV 35).

THE MAGNETRON

24. Description

The magnetron is a valve used for generating UHF oscillations. It has two electrodes; a cylindrical cathode and, coaxial with it, a cylindrical anode. The original magnetron had an undivided anode, but practical modern magnetrons have anodes split into a number of

segments, usually equal in size, and separated by gaps cut parallel to the axis.

An electric field is established between anode and cathode. A magnetic field is applied parallel to the axis of the electrodes, and is as uniform as conditions will permit, over the whole space between them. Because of the magnetic field, the electrons do not move directly from cathode to anode, as in a simple diode, but are constrained to move in curved orbits, as described below. In virtue of this, electrons can be made to arrive at desired segments of the anode in such a manner as to give rise to and sustain high frequency oscillations in tuned circuits suitably connected to the anode segments.

Practical magnetrons are of three main types:-

- (i) The split-anode magnetron operating in the Dynatron (or Habann) mode.
 - (ii) The Megaw split-anode magnetron.
 - (iii) The resonant cavity centimetric magnetron, as widely used in radar.
-
- (i) The Habann oscillator works with relatively high efficiency (30 - 70%) up to about 600 Mc/s. It is not at present in use in service equipments, but a brief description of it is given as an example of how negative resistance effects can arise due to changes in field distribution and in consequence of the electron orbits. (Sec. 26).
 - (ii) The Megaw valve is a low power magnetron, enclosed in a glass envelope of normal size and shape. In contrast to the Habann oscillator, it is a microwave generator. It is not used in British Radar equipments, and it is not dealt with in these notes.
 - (iii) The resonant cavity centimetric magnetron is the only centimetre-wave oscillator so far produced capable of giving pulse powers of the order of megawatts with high efficiency. For operation on centimetre wavelengths the tuned circuits are so small that they are embodied in the anode block of the valve and are not, in general, variable. Fig. 4.22 illustrates the essential features of this type of magnetron.

25. Electron Orbits

The operation of any kind of magnetron is determined largely by the motion of the electrons in the anode-cathode space. The orbits of the electrons depend upon three factors, namely:-

- (i) the electric field due to the anode-cathode potential difference,
- (ii) the applied magnetic field, and
- (iii) the space charge effect, i.e., the forces which the electrons exert on one another.

Early simple theories of magnetron operation which ignored (iii) were based on the calculation of the orbit of a single electron as follows:-

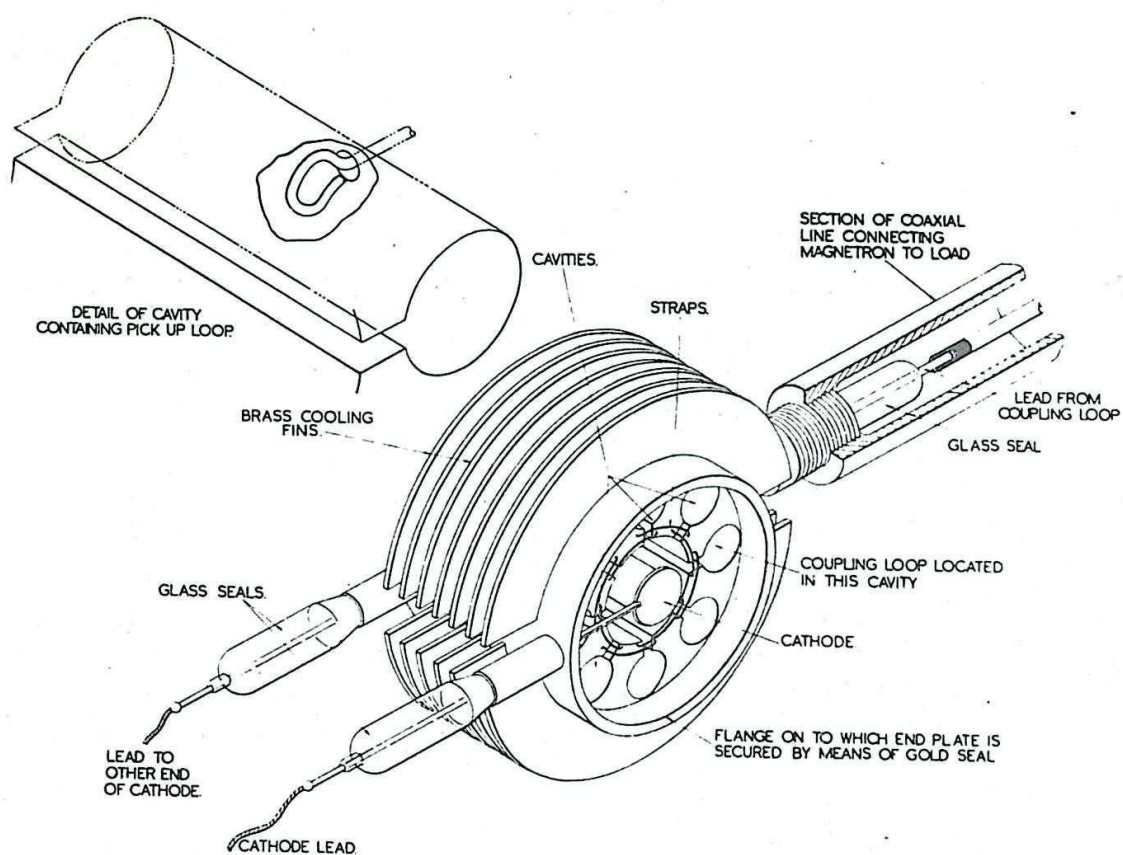


Fig. 422 - Typical hole and slot type magnetron with end plate removed.

Assume that the anode is a uniform cylinder.

Let V be the potential difference between the cathode and any point distant r from the axis.

Let u be the velocity acquired by a electron due to the accelerating effect of the potential difference V .

Denote by suffixes a, k the values of V, u , and r at anode and cathode respectively.

Let m be the mass and e the magnitude of the charge of an electron.

It can be shown that

$$V = V_a \frac{\log r/r_k}{\log r_a/r_k} \dots\dots\dots(1).$$

Fig. 423 shows how V varies with r for the case where $r_k \ll r_a$.

Since the kinetic energy acquired by the electron is $\frac{1}{2} m u^2$, and the work done on the electrons is eV ,

$$\frac{1}{2} m u^2 = eV, \text{ so that}$$

$$u = \sqrt{\frac{2eV}{m}}.$$

In particular,

$$u_a = \sqrt{\frac{2eV_a}{m}} \dots\dots\dots(2).$$

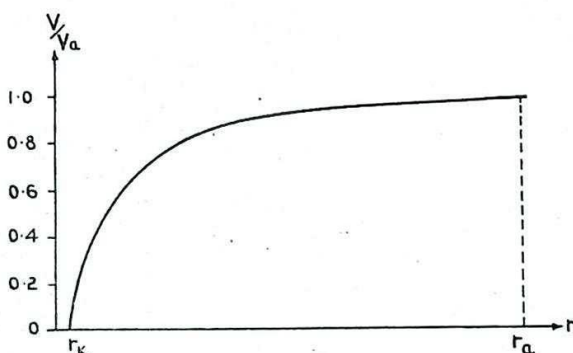


Fig. 423 - Potential distribution in the anode-cathode space.

Fig. 423 shows that, for $r_k \ll r_a$, the potential gradient is steep near the cathode so that $V \approx V_a$ and hence $u \approx u_a$ except within a short distance of the cathode.

So far, only one factor influencing the electron orbit has been discussed, namely the effect of the electric field between anode and cathode. Due to it the electron would move along a radius from cathode to anode, with a velocity which increases at first rapidly and thereafter, very slowly (Fig. 424(a)).

Now if an electron moving with velocity u is subjected to a force F due to a magnetic field, of flux density B , perpendicular to the plane of its motion

$$F = Beu,$$

and is mutually perpendicular to B and to u . This causes the electron to move in a circular orbit, of radius ρ , given by

$$Beu = \frac{mu^2}{\rho} \dots\dots\dots(3).$$

Since, except within the region close to the cathode, u is

approximately constant and equal to u_a , it is valid, as a first approximation, to equate the value of u obtained from equation (3) to that for u_a in equation (2).

This gives

$$B = \sqrt{\frac{2m V_a}{e \rho^2}} \quad ; \quad \text{or} \quad \rho = \sqrt{\frac{2m V_a}{e B^2}}.$$

Thus every electron leaving the cathode region commences to describe a path which is approximately circular. Whether or not the path is completed depends upon the values of V_a and B . Fig. 424(b) shows the case where ρ is very large; Fig. 424(c) where, by increasing B , ρ has been reduced to $\frac{r_a}{2}$, so that the electron just skims

the anode, and Fig. 424(d) shows ρ made small, by suitably increasing B . For the case of Fig. 424(b), all electrons would reach the anode; whereas for that of Fig. 424(d) none would do so; i.e., the valve current would be cut off. Fig. 424(c) represents the case where B is just large enough to prevent the electrons from reaching the anode; i.e., the magnetron current is just cut off. The value of B required to achieve this is called the cut-off value B_c and is obtained by putting $\rho = \frac{r_a}{2}$ in the expression for B above. This

gives

$$B_c = \sqrt{\frac{8m V_a}{e r_a^2}} \dots\dots\dots (4).$$

B OUT OF PAPER

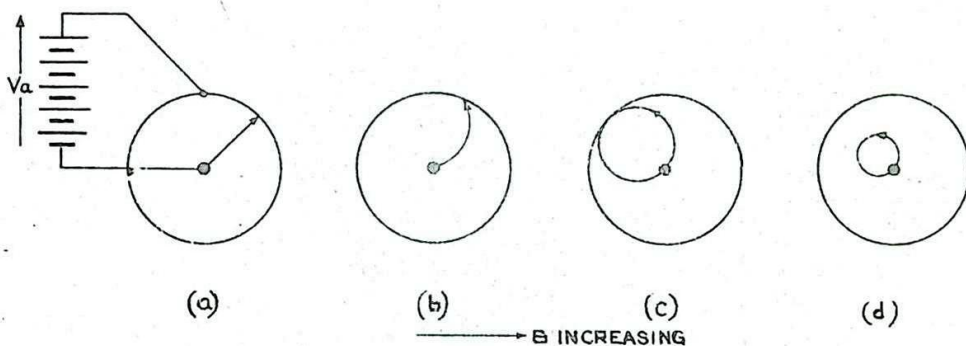


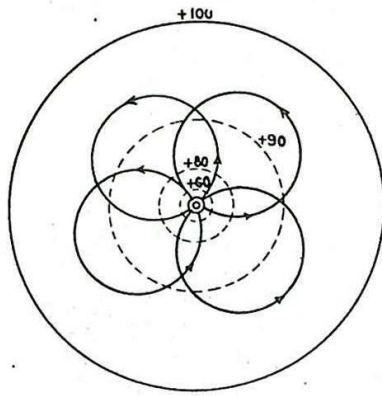
Fig. 424 - Single electron orbits with fixed V_a and varying B .

(Actually, the orbits cannot be true circles, for the assumption that u is constant is only an approximation. The true orbit curves more sharply near to the cathode, and is roughly epicycloidal in shape).

Fig. 425 indicates the circular orbits of electrons, for V_a and B constant. It is seen that their centres lie on an equipotential surface coaxial with the cathode; and that a similar coaxial surface marks the space charge boundary.

The existence of the space charge modifies the above theory. Electrons on their way out from the cathode are repelled by those nearer the anode, and are unable to execute the orbits so far described. Also, electrons which are attempting to return to the cathode are prevented from so doing by those nearer the cathode. It is more probable that these outer electrons, since they cannot fall back into the cathode, move in circular orbits around it.

B OUT OF PAPER



----- EQUIPOTENTIAL SURFACES.
 ———— ELECTRON ORBITS

Fig. 425 - Electron orbits for a cut-off magnetron with V_a and B constant and uniform.

B OUT OF PAPER

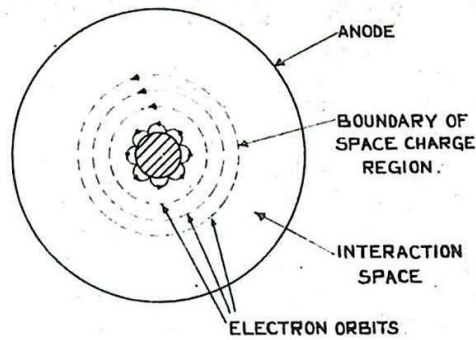


Fig. 426 - Electron orbits as modified by the space charge.

Fig. 426 indicates the electron paths obtaining when a steady anode voltage has been established, B being larger than the cut-off value. It is useful to regard the space charge region as composed of a number of thin coaxial cylindrical shells each containing a large number of electrons which revolve with the same angular velocity. The angular velocity depends upon V_a , B and the radius r of the shell. The greater r , the larger is the angular velocity of the electrons.

Up to the present, only the orbital conditions obtaining when V_a is fixed have been considered, Figs. 424 and 425 indicating the orbits when space-charge effects are ignored and Fig. 426 when they are not. It may be shown that whichever orbit is assumed, much the same result is obtained, this result according well with the practical performance of the Habann oscillator. This oscillator will be examined from both points of view and thus the effect upon the orbits of varying V_a will be introduced.

26. The Split-Anode Magnetron Operating in the Dynatron Mode (Habann Oscillator)

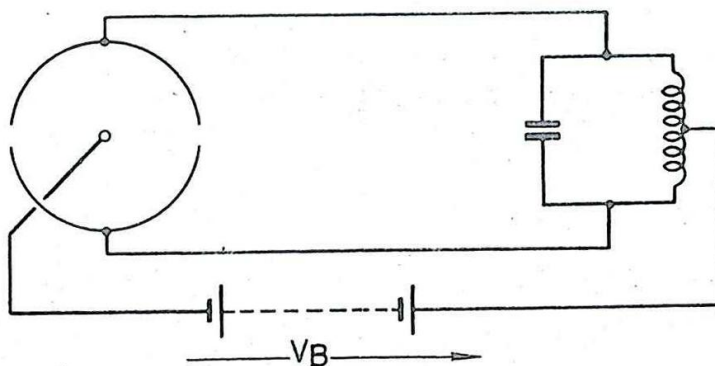


Fig. 427 - The two-segment magnetron.

This mode occurs at frequencies such that the period of oscillation is long compared with the time of transit of an electron from cathode to anode, and normally arises for waves longer than about 50 cms. In such circumstances it is valid to assume that, during their passage from cathode to anode, the electrons are in a field that is approximately steady.

Consider a two-segment magnetron, connected as shown in Fig. 427. Suppose that initially each segment is at a potential V_0 above that of the cathode, B being larger than B_c for this particular value of the anode voltage. Then Fig. 428(a) shows the equipotential lines and electron orbits under this circumstance, space-charge effects being ignored. Fig. 428(b) gives similar information, but shows the concentric circular orbits which appear reasonable if the space charge is taken into account. In this latter case, the orbits follow the equipotential lines, whilst in the former, the orbit centres lie on an equipotential line.

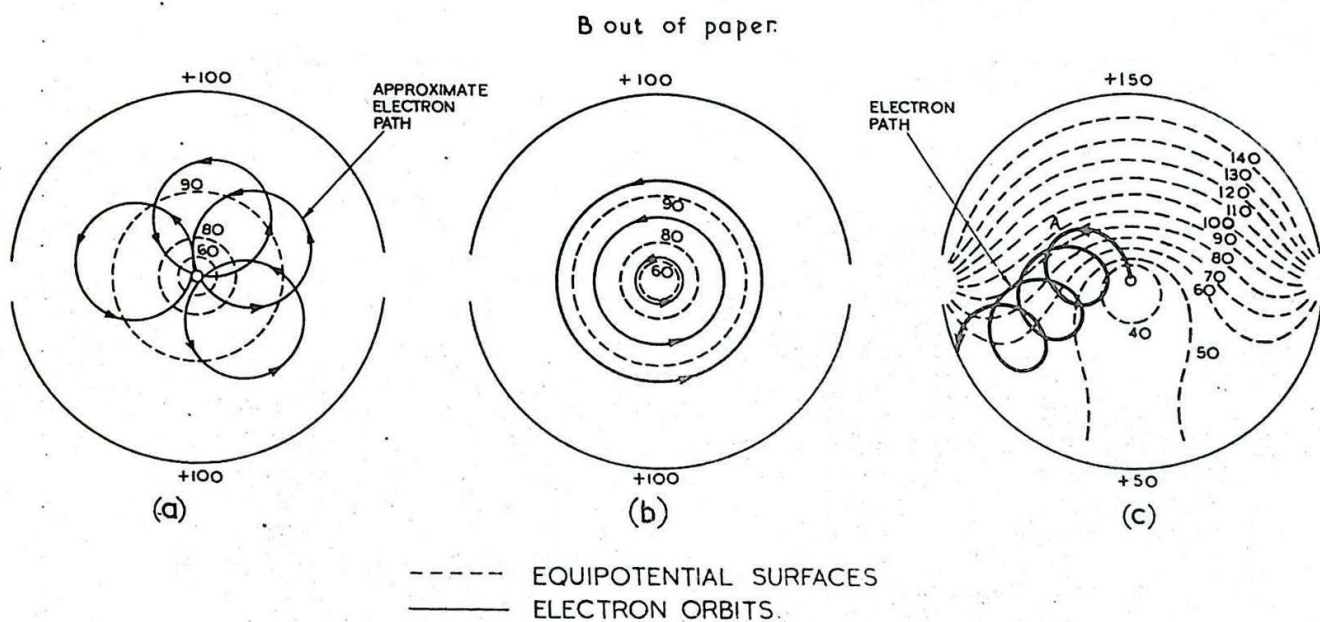


Fig. 428 - Orbits of a single electron in a split-anode magnetron.

Now let the potential of the upper segment be raised to $(V_0 + V_1)$, whilst that of the lower segment is reduced to $(V_0 - V_1)$. The field configurations and equipotential distributions are consequently changed as indicated in Fig. 428(c). In the absence of space charge effects, the electrons still move in such a manner that the orbit "centre line" is always approximately at right angles to the electric field; i.e., this line still corresponds to an equipotential line. Fig. 428(c) shows an electron orbit for such a case, and if the similar orbits of many electrons, leaving the cathode at all parts of its surface, be considered, it is readily seen that more electrons will arrive at the lower segment than at the upper.

If now the effect of space charge be considered, an electron at A (Fig. 428(c)) will be prevented from moving in towards the cathode by other electrons and will therefore be constrained to move on a path approximating to the envelope of the orbit such as that shown

in the figure; i.e., the orbit itself tends to follow an equipotential line. Electrons in other portions of the anode-cathode space similarly tend to follow equipotential lines. Hence the introduction of the effect of the space charge, whilst it modifies considerably the individual electron orbits, in no way alters the conclusion, viz, that more electrons arrive at the lower potential segment than at the higher. Further, it can be shown that if the circular orbits assumed in the presence of space charge (and fixed V_a) be postulated, the cut-off value of B is given by

$$B_0 = \sqrt{\frac{8m V_a}{e \left(r_a - \frac{r_k^2}{r_a}\right)^2}} :$$

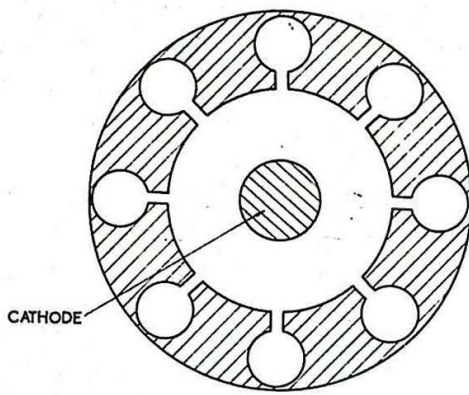
and if $r_a \gg r_k$ this reduces to equation (4), which was obtained on the assumption of an entirely different orbit. This fortuitous agreement has been largely responsible for the continued neglect of consideration of space-charge effects as in the early magnetron theories, but, although such omission leads to negligible errors, except in the orbit concept, for the Habann oscillator, it gives erroneous results for the resonant cavity magnetron considered in Secs. 27 - 42.

Before proceeding to this case, it is instructive to consider how, in the Habann oscillator, the conditions required for the maintenance of oscillation are fulfilled. We have seen that more electrons arrive at the lower potential segment than at the higher. That is, referring to Fig. 427, the existing negative charge on the lower condenser plate is increased because of this, by an amount (say) ΔQ . If the instantaneous value of the RF voltage on the condenser is v , its RF energy is thus increased by $\frac{1}{2} v \Delta Q$. By arranging that

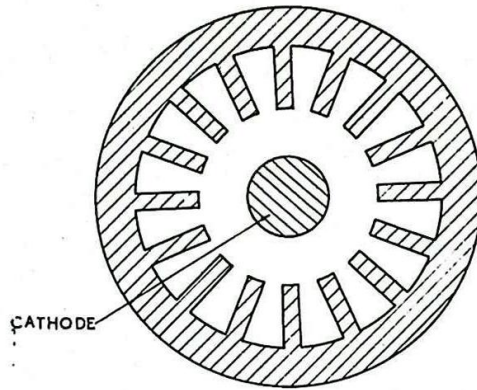
most electrons arrive at the lower segment when C is fully charged (i.e. when v has its maximum value) the greatest possible increment of energy is given to the tuned circuit. Half a periodic time later the upper plate of the condenser will be negative with respect to the lower; and at this instant the majority of the electrons will be arriving at the upper plate. Thus the oscillatory circuit receives increments of energy due to the arrival of electrons at the right time. Hence dissipation in the tuned circuit (or in whatever load is coupled to this circuit) may be made up by the magnetron. Since the current through each segment is antiphase to its RF voltage, the magnetron behaves as a negative resistance element; this accounts for the name "Dynatron mode" often used to describe this kind of magnetron operation. This mode can be efficient only if the electron cathode-to-anode transit time is negligible compared with the periodic time, for only then will each increment of charge be received when the corresponding segment voltage is at a minimum.

27. The Resonant Cavity Centimetric Magnetron

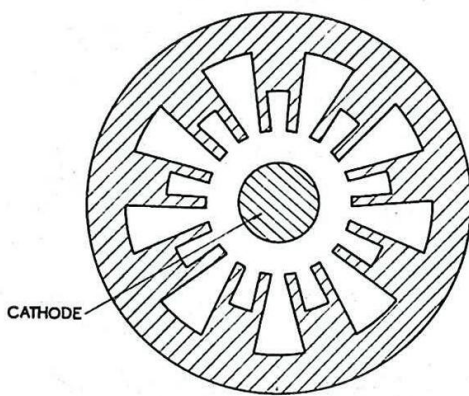
The construction of a typical magnetron of this type is indicated in Fig. 422. The tuned circuits are holes and slots cut in a solid beryllium-copper block; Fig. 429 illustrates some common anode block constructions. Considering one hole and slot apart from the remainder, we may regard the slot as a condenser, since most of the electric field is located at it, and the hole as, roughly, a single turn coil. Thus a single hole and slot constitute a resonant circuit. The equivalent circuit of the magnetron as a whole is more complicated than this, not only because there are many cavities, but also because they are coupled one to another in a complicated manner. Firstly there is mutual induction between the holes, since magnetic flux emerging from the top of one hole passes through an adjacent one to form a closed loop, (see Fig. 430(a)). Also, there



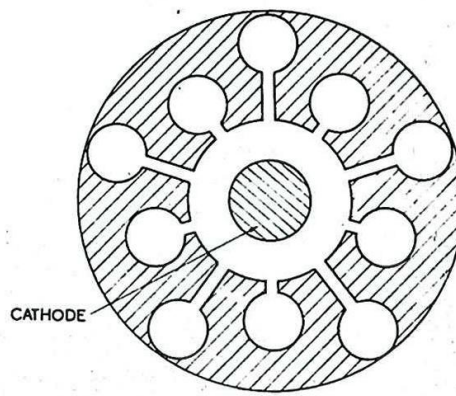
(a) HOLE & SLOT TYPE OF BLOCK.



(b) VANE TYPE OF BLOCK.



(c) "RISING SUN" VANE TYPE OF BLOCK.



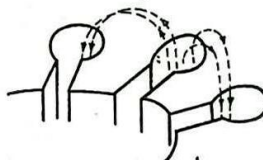
(d) "RISING SUN" HOLE & SLOT TYPE OF BLOCK.

Fig. 4.29 - Types of anode block construction in use.

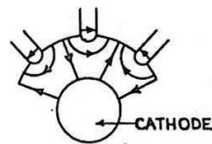
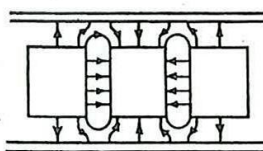
are capacitances to the end plates and to the cathode, giving rise to capacitive coupling between the cavities. Coupling also occurs through the medium of the space charge itself. Since an N-segment magnetron may be regarded as having a simple equivalent circuit given roughly by placing the N-segments in parallel, the resultant equivalent lumped circuit has a capacitance N times and an inductance

$\frac{1}{N}$ th of that of a single

cavity. Thus the operating frequency would appear to be that of a single cavity. In practice, whilst the cavity dimensions are the most important single factor in determining the frequency of operation, they are influenced materially by the space-charge cloud, as well as by the values of the applied voltage and the magnetic flux density. Hence it is necessary to consider in some detail the behaviour of the space charge.



(a) MAGNETIC (INDUCTIVE) COUPLING BETWEEN HOLE AND ITS NEIGHBOURS.



ELECTRIC (CAPACITIVE) COUPLING BETWEEN SEGMENTS AND END-PLATES (LEFT-HAND SKETCH) OR BETWEEN SEGMENTS AND CATHODE (RIGHT-HAND SKETCH).

(b) COUPLING BETWEEN SEGMENTS.

Fig. 430 - Coupling between segments.

28. The Single Stream Steady State

The following account is now generally accepted as describing what occurs in the anode-cathode space of a resonant cavity magnetron. It is based on the concept of the Single Stream Steady State. Each electron is assumed, if V_a is constant, to be moving in a circular orbit about a point on the axis as centre. If V_a is varied the electrons change their orbital radii accordingly. It is assumed that the changes of anode voltage occur so relatively slowly that the electrons immediately move into the orbit corresponding to the new value of V_a , so that any instantaneous condition is virtually a "steady state", i.e. one obtaining with fixed V_a . Since magnetrons are pulse-operated, the validity of this assumption might appear doubtful; but present pulse technique cannot produce modulating pulses rising to their full value in less than about $0.02 \mu\text{S}$. Electron orbits can automatically adapt themselves to changes of this order, so that a succession of "steady" states occurs as the pulse is applied. Hence at any point in the space-charge region the electrons are, at any instant during the change, either all going towards or all going away from the cathode. This is the distinction between Single Stream and Double Stream steady states, for under the double stream condition some electrons may be moving through the point towards the anode and others towards the cathode. Analysis has shown that the latter condition would obtain if the pulse rise were instantaneous. Since we have not yet approached sufficiently near to this condition for the single stream state to be invalidated, the possibility of a double stream steady state will hereafter be disregarded. (It may be noted that the early concept of electrons hurtling among one another in an interweaving of epicycloids implied a double stream state). Also, with the rates of rise of the modulating pulse encountered in practice, it is not necessary to differentiate between CW and pulse operation.

29. Behaviour of the Space-Charge Cloud when the Anode Voltage Varies

So far, discussion has been limited to the operation of the magnetron with V_a fixed. The next step is to consider what happens

to the space-charge cloud when oscillations of the tuned circuits (cavities) cause the anode voltage to vary.

Any disturbances at the anode, resulting in oscillations, sets up a voltage pattern upon it. This pattern can be analysed into component waves revolving at different angular rates around the anode surface.

One of these component waves roughly in step with a particular "electron shell" (see Sec. 25) may cause considerable distortion of its cylindrical form. This distortion in turn affects the motion of other electrons. It can be shown that electrons, between this shell and the anode, and which, with constant V_a , move faster than those in it, are slowed up when the space charge cloud is deformed, and that those nearer the cathode have their angular velocities increased. Thus the electrons outside the "in-step" shell give up energy whilst those inside gain energy. If the outer electrons give up more energy than is acquired by those nearer the cathode there will be a net loss of energy by the space charge, and if this energy can be taken up by the tuned circuits a means of sustaining oscillations is provided. This condition, although necessary, is not sufficient, for the energy of the space charge must be replenished from the DC source. This can be achieved only if the deformation of the space-charge cloud is sufficiently great to cause electrons, in sufficient numbers, actually to reach the anode, and this clearly depends upon the value of V_a . (It depends also upon the nature of the voltage pattern around the anode.)

Before considering the factors mentioned above in greater detail, a rough picture of the part played by the space charge, and of the operation of the magnetron generally, may be obtained with the aid of the following analogy.

30. The Electrometer Analogy

Consider a light paddle-shaped vane (Fig. 431(a)) supported by a torsion head and deflected by the charged surfaces ABCD. The vane takes up a position of equilibrium determined by the deflecting torque due to the system of charges, and the restoring torque due to the suspension. (This is a form of what is generally known as a quadrant electrometer).

If the constraint on the vane is removed and the cylindrical outer surface rotated, the vane will keep step with the rotation. Alternatively, the same result could be achieved by the use of a commutator in the leads to the segments, arranged so as to reverse the charges periodically. Such a system can be regarded as an elementary motor and could be made to perform work.

Conversely if the rotary vane in the arrangement shown in Fig. 431(b) were driven by an external motor, charges would be induced on the segments as indicated, and would vary periodically. The resulting voltages would cause currents to flow in the external circuit. Such a system is clearly a simple generator, and could be made to supply power to a load.

In the magnetron there is a state of affairs resembling this in several respects. The segments of the electrometer correspond to the anode segments of the magnetron, and the charged vane to the rotating space charge. Figs. 432(a) and (b) show the space charge distorted due to the segment potentials, for the 4-segment and 8-segment cases. It will be noted that in neither case shown could oscillation be maintained, for no electrons reach the anode, and

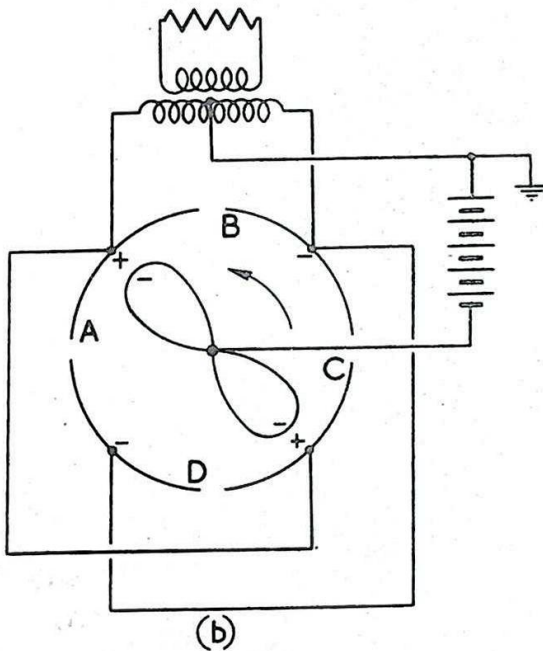
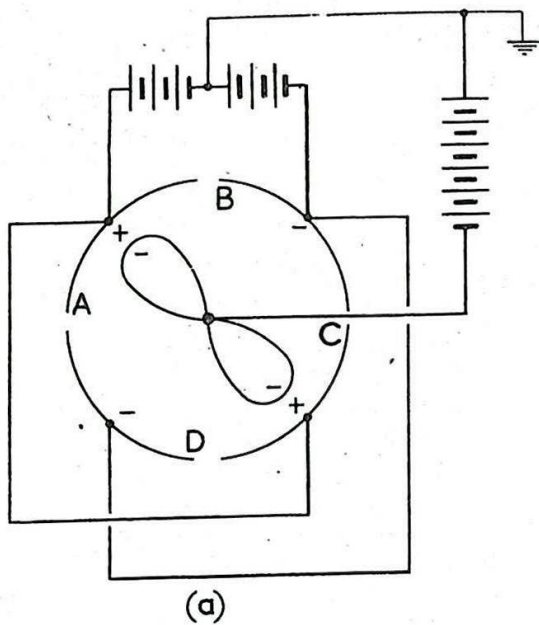


Fig. 431 - The electrometer analogy.

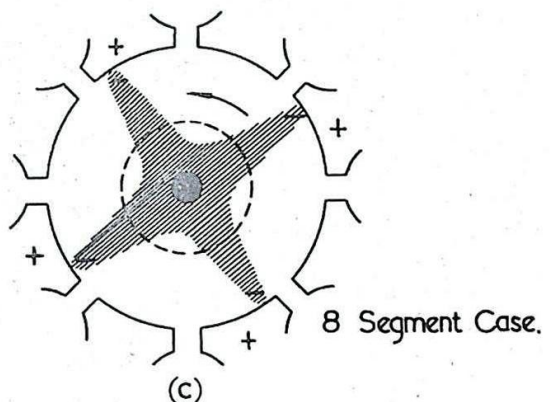
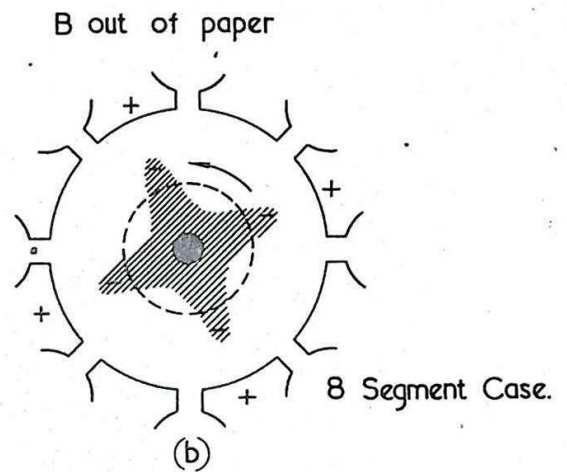
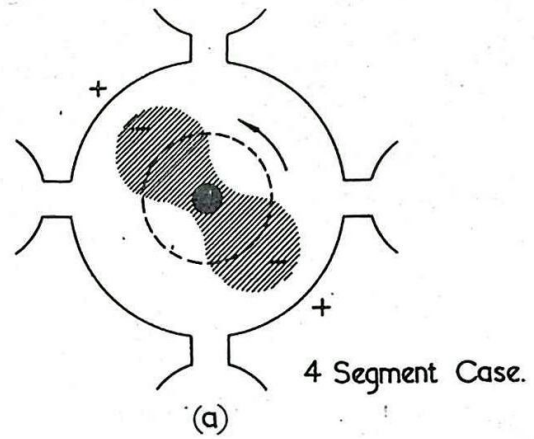


Fig. 432 - Space charge deformation in the magnetron.

and hence no energy is being supplied to the space charge. Fig. 432(o) shows the cloud sufficiently distorted for this to occur. Under this condition there will be an inwards radial flow of current in the magnetron (corresponding to the outward flow of electrons). This current flows under the influence of an external magnetic field, and the space charge is accordingly constrained to revolve. (In the case shown, the magnetic flux is out of the paper, and the consequent motion is counterclockwise). Thus we have an effect equivalent to the driving motor in the analogy.

31. Modes of Oscillation

It is next necessary to consider in further detail the voltage pattern on the anode referred to in Sec. 29. Common anode block arrangements have been shown in Fig. 429. Consider that of Fig. 429 (a). If a slot, regarded as the condenser of the equivalent tuned circuit, be charged up in some way, the cavity will oscillate at its natural frequency. Now such a cavity has many possible modes of oscillation, and in a magnetron the existence of other cavities all coupled by magnetic and electric fields and by the space charge, complicates the problem considerably.

The simplest and commonest mode is one in which at any instant alternate segments are at equal and opposite RF potentials, i.e. there is a phase difference of 180° (π radians) between adjacent segments. This mode is commonly called the π -mode. The instantaneous potential of any segment varies sinusoidally with time. Fig. 433(a) shows the instantaneous voltage distribution around the anode at intervals of one eighth of a periodic time T , for an eight-segment magnetron, the segments being drawn for convenience in a straight line. Whereas the time variation of voltage at any point on the anode is sinusoidal, the voltage distribution plotted against distance round the anode is not. That is, the voltage pattern is a sinusoidal time variation, of amplitude varying periodically with the angular distance; i.e., it is a standing wave. Further this standing wave can be analysed into a series of spatial harmonic components, each of which can be regarded as consisting of two identical travelling waves moving round the anode in opposite directions. Thus, if we take the π -mode here considered, and restrict our attention to the travelling waves comprising the primary or fundamental spatial harmonic only, we shall obtain the results shown in Fig. 434(a). If all the spatial harmonics were included, the waveforms of Fig. 433(a) would result.

The primary component in the standing wave so far considered would have 4 complete repeats around the anode. This is known as its mode number, n . Thus the π -mode in an 8 segment magnetron is also the mode for which $n = 4$; whereas in a 10-segment magnetron the π -mode would be the one for which $n = 5$, since there would then be 5 complete repeats around the anode.

n may have any value, i.e., any phase difference may exist between adjacent segments, provided only that the total phase change round the block is a multiple of 360° . Then each component travelling wave will have a number of repeats n_1 given by $n_1 = mN + n$ where N is the number of segments and m any integer (positive or negative, a negative answer corresponding to a component in the opposite direction). Thus if we take the case so far considered, where $n = 4$ and $N = 8$, $n_1 = 4$ (for $m = 0$), 8 (for $m = 1$), 12, 16, etc., also $n_1 = -4, -8, -12$ etc. for negative values of m . The component for which $n_1 = -4$ is called the primary reverse component, and by combining this with the primary forward component ($n_1 = 4$) we have obtained the resultant curve of Fig. 434(a). Adding in also all the other travelling components would have produced Fig. 434(a).

In a similar manner the Figs. 434(b), (c) and (d) have been produced, for the cases $n = 3$, $n = 2$, $n = 1$ respectively. For putting $n = 3$, $N = 8$, $m = -1$ we obtain $n_1 = -5$ for the primary reverse component. These are shown in Fig. 434(b), which leads to Fig. 434(b) if these and all the other travelling components are added. The reader is left to confirm for himself that the correct values for the primary reverse modes have been taken in Fig. 434(c) and 434(d), and that these results would lead to the anode waveforms shown in Figs. 434(c) and 434(d).

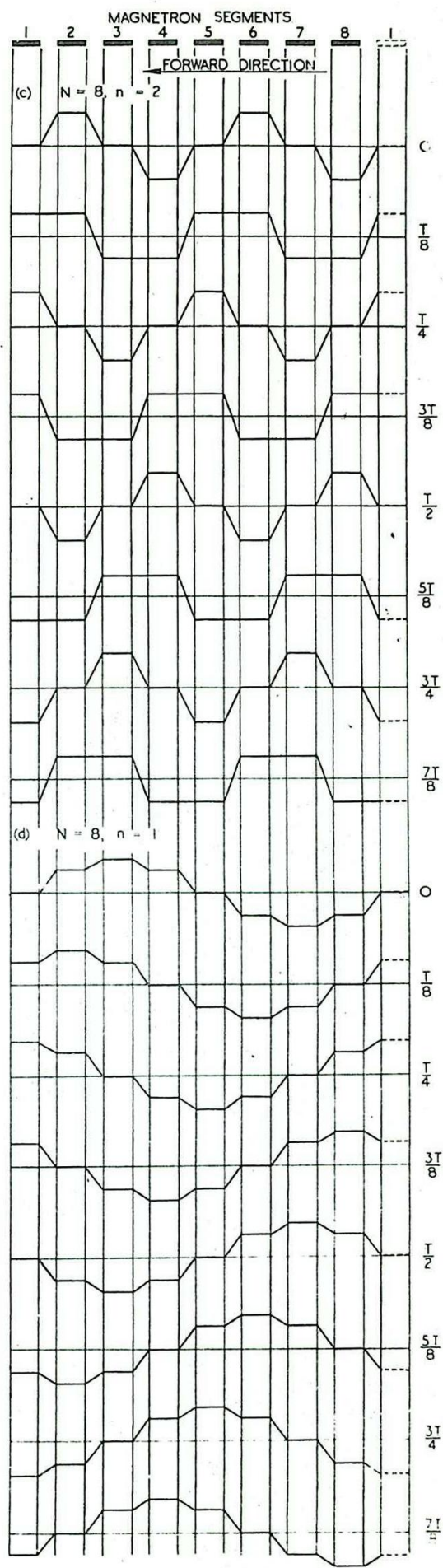
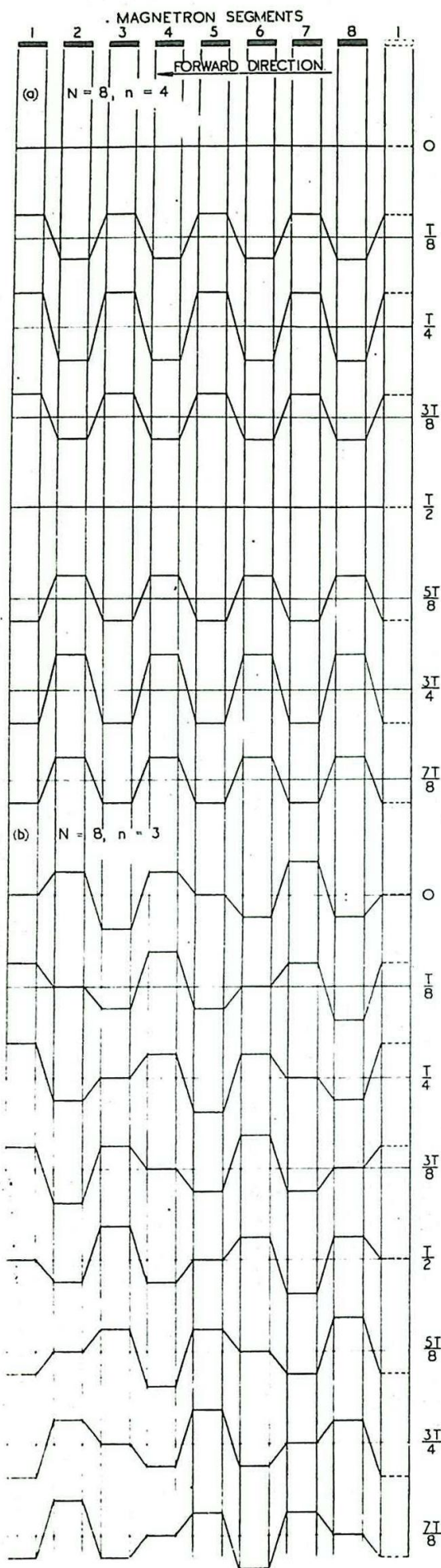


Fig. 433 - Standing waves of voltage on a magnetron anode.

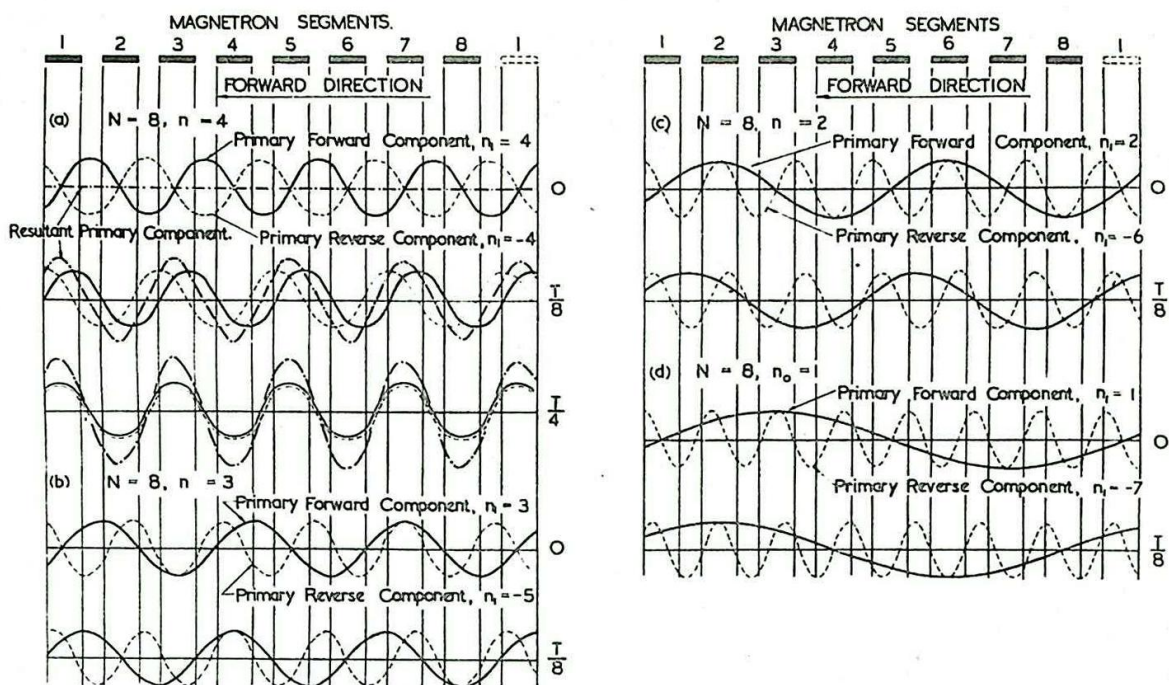


Fig. 434 - Primary components of standing waves of voltage on an 8-segment magnetron anode.

Since an apparently infinite number of modes appears, from the foregoing, to be possible, it is next necessary to consider how it is that a given magnetron can be made to operate in (say) the π -mode, as is most usual.

32. Initiation of Oscillation - The Energy Transfer Criterion

It has already been stated (Section 29) that variations in the anode potential cause the electrons in the space-charge cloud to be distorted in their orbits, and that an electron shell roughly in step with a component of the anode waveform is markedly affected. Now, since the farther the electrons are away from the cathode, the faster they move in their circular (constant V_a) orbits, a sudden "pulling-out" towards the anode of some electrons will cause those nearer the anode to pile up on those drawn out towards it. Thus, not only will this lead to the "star-fish" shape of the space-charge cloud indicated in Fig. 432(b), but it will cause a retardation of the outer electrons, which will tend to assume approximately the same angular velocity as the component of the anode waveform considered. Thus these outer electrons will give up energy. Electrons nearer the cathode than those in the shell most affected will increase in angular velocity, and hence acquire additional energy. For oscillations to be initiated the energy given up by the space-charge must exceed that taken from it. This requires that an adequate proportion of the space charge must lie outside the shell which is in synchronism with the anode waveform component. It can be shown that whereas the distribution of the space charge clearly depends on V_a (assumed constant) the radius of an electron shell in which the electrons travel with a definite velocity does not. For a mode of oscillation with frequency f and mode-number n the synchronous velocity is a

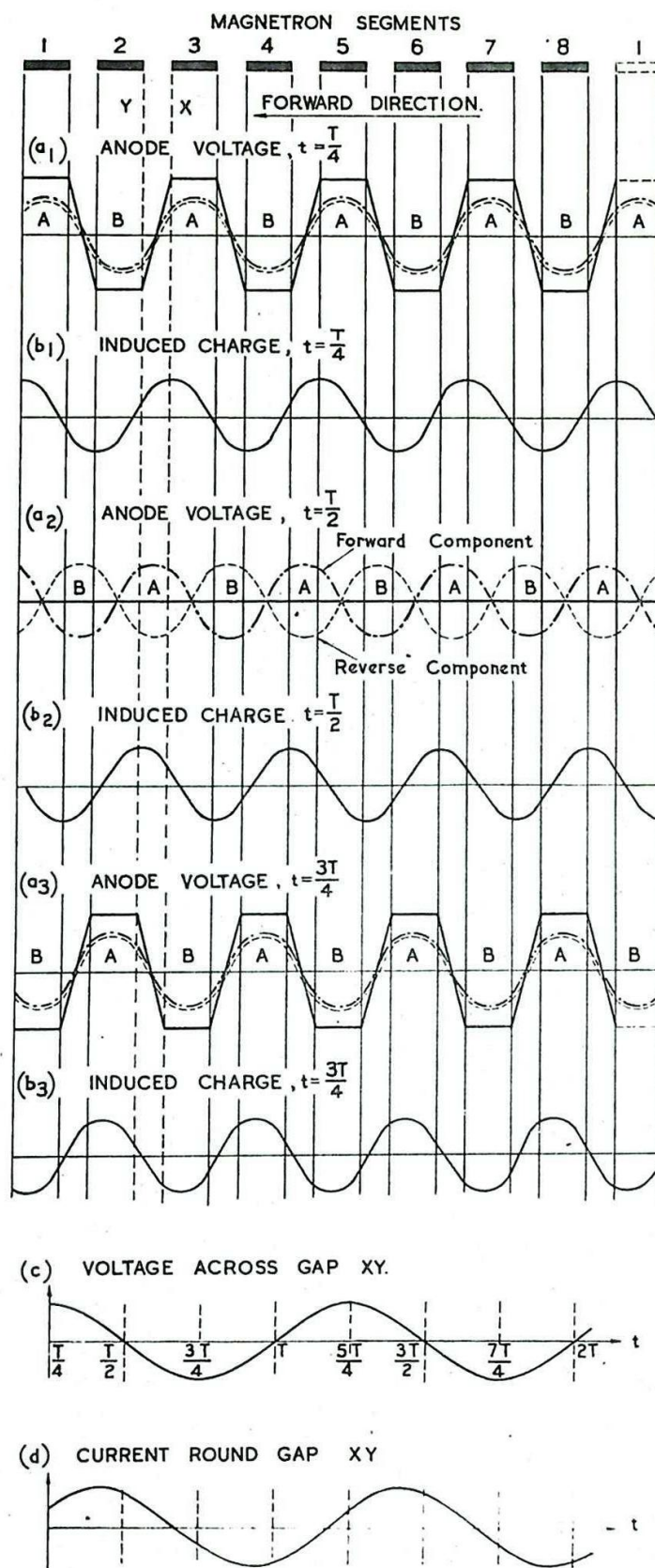


Fig. 4.5 - Distribution of anode voltage and charge under oscillatory conditions.

function of f/n . The radius of the shell moving with this velocity depends on B and r_k but not on V_a . Also, the larger n , the smaller is this radius.

The magnetron must therefore be operated with V_a large enough so that the radius of the space-charge cloud is greater than the radius of the synchronous shell by an amount sufficient to ensure the transfer of energy from the space charge to the cavities. This condition will be termed the Energy Transfer Criterion. The value of V_a required to satisfy this criterion will be called the "Energy Transfer Potential". Since the radius of the critical shell is smaller the larger the value of n , to operate the magnetron in a high order mode necessitates a smaller anode voltage than operating it in a low order mode; also, if V_a be steadily increased, the energy transfer criterion is satisfied successively for lower and lower order modes.

All components of the anode waveform affect the electron cloud to some extent. But only a forward component, i.e. one moving in the same direction as the space charge as a whole, can be in synchronism with an electron shell. Reverse components have little effect on the general motion of the space charge as a whole. Further, it can be shown that the potential V_1 due to a wave component n_1 , at a point within the space charge, distant r from the axis, increases with $(\frac{r}{r_a})^{n_1}$ so that, except near to the anode, the contributions to V made by low values of n_1 will be larger than those made by high values of n_1 . Hence it appears that, from these considerations alone, the primary forward mode would cause the greatest distortion of the space-charge cloud.

It remains to be shown that the energy available when the energy transfer criterion is satisfied can actually enter the cavities; i.e., that the power flux is directed into them. In order to see this, consider again the 8-segment magnetron, operating in the $n = 4$ mode. We shall again take only the primary components, as shown in Fig. 435, in which the components of Fig. 434(a) are repeated. Also we shall assume that the primary forward component is the major effective component for the initiation of oscillation. In Fig. 436, the conditions in the magnetron at time $t = \frac{T}{4}$ are

illustrated. The line OA is drawn through the instantaneous crest of the primary forward component of the anode waveform; OB through the trough. It is shown in the full mathematical theory of the magnetron that the aggregation of electrons takes place in such a manner that the bulges in the distorted space-charge cloud are always ahead of crest lines such as OA. At points on the anode opposite to each bulge there will be, by induction, a local increase in the positive charge on the anode, and, correspondingly, there will be a local reduction in the

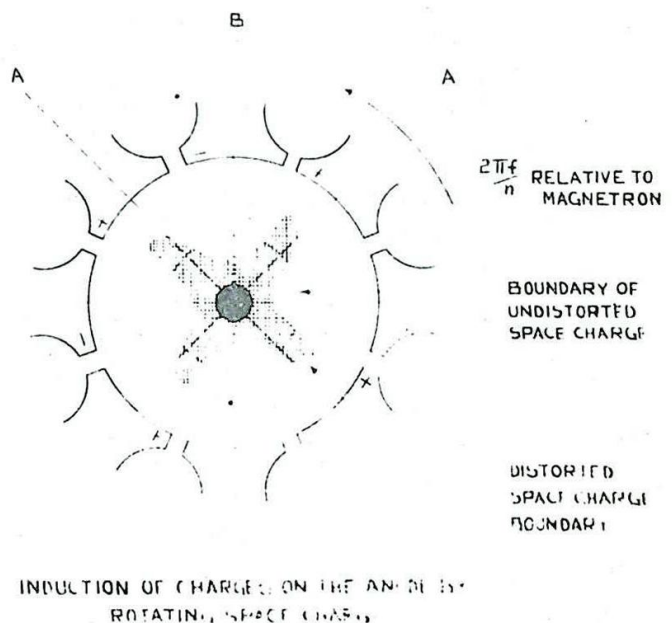


Fig. 436 - Induction of charges on the anode by rotating space charge.

charge on the anode opposite each trough in the space-charge cloud. Thus, superposed upon the positive charge on the anode due to the applied voltage, there will be a variation of charge corresponding to the deformation of the space charge. This variation is indicated on Fig. 436 by plus and minus signs.

The spatial distribution of induced charges on the anode must be periodic, and in the case taken will have four periods round the anode. In the diagram (Fig. 435) only the fundamental component of this distribution is shown, as this is the only component which contributes appreciably to the output power. As the primary forward component of voltage moves round the anode, the space charge deformation, and therefore the variation of charge on the anode, moves with it. The distribution of charge at later instants is shown in Figs. 435(b2) and (b3).

Consider a particular gap XY. The spatial distribution of the anode voltage at successive instants is shown in Figs. 435(a1), (a2) and (a3), and the corresponding time-variation of the voltage across the gap XY is shown in Fig. 435(c). The relation between the component of the charge distribution and the potential of X relative to Y may now be deduced. Since the charge distribution is moving round the anode in an anticlockwise direction, there is a component of current flow at every point on the anode, proportional to the corresponding component of instantaneous charge density at the point. Hence the graphs of Fig. 435(b1), (b2) and (b3) may be interpreted as current distribution round the anode, instant by instant, a current flowing in an anticlockwise direction being reckoned as positive. The manner in which the current flowing from X to Y, i.e. around a cavity, varies with time may be deduced from these graphs, and is plotted in Fig. 435(d).

Fig. 437 shows the particular cavity considered. When X is positive with respect to Y the direction of the electric field across the slot is as shown by the arrow E. Hence if electric field strength is regarded as positive in the direction of the arrow, Fig. 435(c) may be usefully taken as a graph of electric field intensity. When the current round the cavity (Fig. 437) is positive the direction of the magnetic field is as shown by H (out of the paper). Hence if magnetic field intensity is regarded as positive in this direction, Fig. 435(d) may be taken as a graph of magnetic field intensity.

If the cavity were resonating without gain or loss of energy, the time-phase relation between E and H would be a quadrature one. But superposed on the magnetic field due to resonance of the cavity, there is in the magnetron an additional

magnetic field produced by the mechanism described above. Comparison of the graphs of Figs. 435(c) and (d) shows that the phase of H differs from that of E by a phase angle less than $\pi/2$, so that H has components both in quadrature and in phase with E. The former indicates that the electronic mechanism of the magnetron modifies the resonating field in the cavity and hence affects the frequency of oscillation. It is one reason why resonant frequencies measured

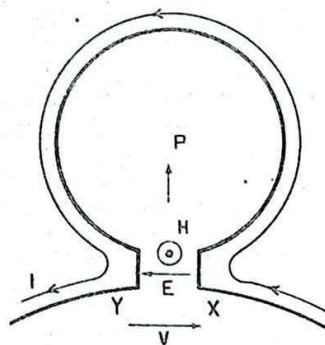


Fig. 437 - Power flux into cavity.

with the cold block differ appreciably from those obtained when the magnetron is actually in use. It is also a reason why the magnetron frequency is sensitive to changes in anode voltage.

The in-phase component indicates that there is a flow of energy, and the direction of the power flux P is obtained from the usual consideration, i.e. that E , H and P form a right handed system. (see Fig. 438). Thus, in Fig. 437, the power flux is directed into the cavity. It is thus possible, provided V_a has the required value, with due regard to n , for energy to be transferred from the space charge to the resonating cavities.

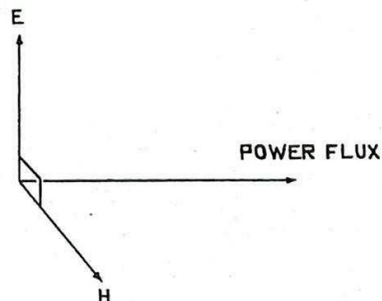


Fig. 438 - Relative dispositions of the electric, magnetic and power flux vectors.

It should be stated that the precise behaviour of the electrons in the shell most affected by the anode waveform, and also the exact location of this shell, is not yet firmly established. There are some differences in the views held by various research workers. It is, however, agreed that, although a certain shell (or shells) may have greater importance than others in the operation of the magnetron, the essential factor is the behaviour of the space-charge cloud as a whole. Oscillations are possible only if the cloud gives up more energy than it gains; and this condition can be achieved, in a particular mode, only if V_a has sufficiently high value; that is, if the energy transfer potential is exceeded.

33. Maintenance of Oscillation: The Energy Supply Criterion

As already observed, oscillations can be maintained only if the energy transferred from the space-charge cloud to the cavity is replenished from the HT supply; i.e. some electrons in the space-charge cloud must have sufficient energy to reach the anode.

This condition, which will be termed the Energy Supply Criterion, may be determined mathematically, and it is found that, as in the case of the energy transfer criterion, V_a must be greater than a critical value, called the Energy Supply Potential (or Threshold Potential). This potential also increases with the ratio $\frac{f}{n}$, so that as V_a is increased from zero the energy supply criterion is satisfied first for the higher order modes and successively for those of lower order. For given values of f , n and B the energy transfer and supply potentials are not the same.

34. The Two Criteria Combined

It follows from Secs. 32 and 33 that for a magnetron to oscillate continuously in a given mode its anode supply voltage must exceed both the energy transfer and the energy supply potentials.

It has been previously stated (Sec. 26) that B_c , the cut-off value of B for a given V_a , is given by

$$B_c = \sqrt{\frac{8n V_a}{e \left(r_a - \frac{r_k^2}{r_a} \right)^2}},$$

so that $V_a \propto B_0^2$. Hence the curve showing the variation of B_0 with V_a is a parabola, indicated by OS, Fig. 439. Further, it has been mentioned already that the energy supply potential depends upon both B and n . It can be shown that for given values of n the graphs of the energy supply potential against B yield straight lines which are tangential to the cut-off parabola, as shown. (The points of contact are calculable, and are nearer the origin the larger is the value of n). Corresponding graphs showing how the energy transfer potential varies with n are shown by the dotted lines in Fig. 439.

For a given V_a and B , a point may be plotted on the composite graph representing the operating conditions. For example, Fig. 439 shows that, for the point P,

- (i) V_a is insufficient to cause all electrons to reach the anode; V_a would have to be increased to the ordinate of S for this to happen.
- (ii) that the energy transfer potential for $n = 5$ is exceeded
- (iii) that the energy supply potential for $n = 5$ is exceeded also.

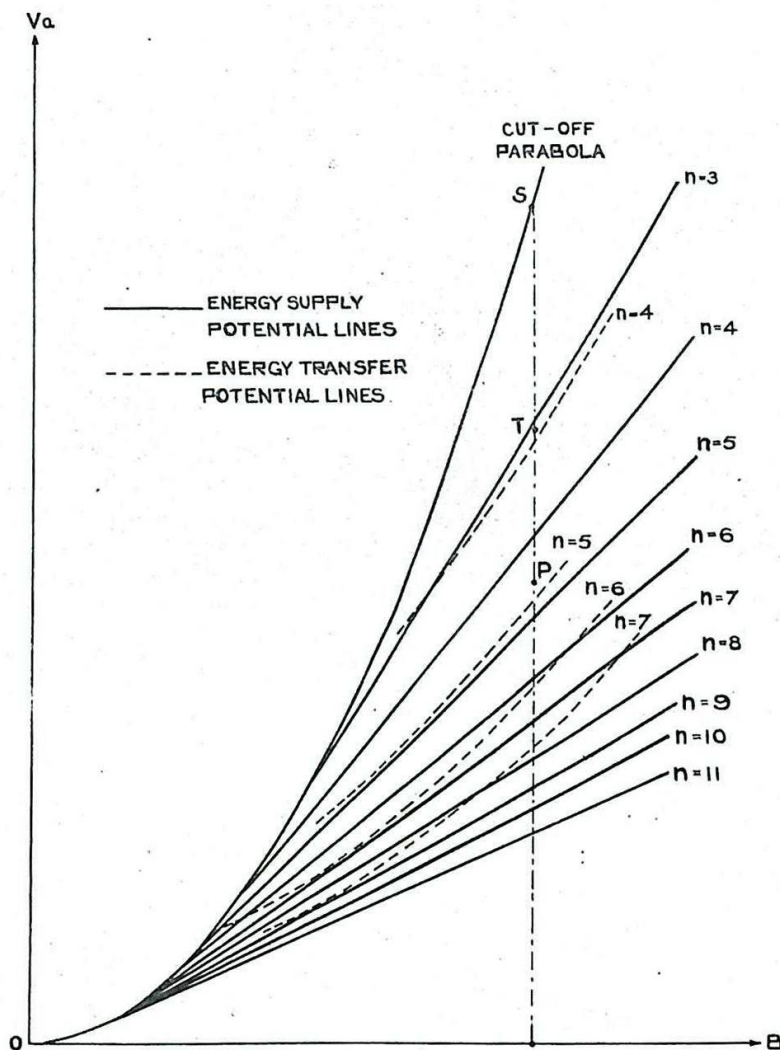


Fig. 439 - Operating conditions for an unstrapped magnetron.

(ii) and (iii) combined given the information that oscillation in the $n = 5$ mode is possible. The diagram also shows that oscillation in any lower order mode is not. With the working point at P, however, not only are the conditions for oscillation in the mode $n = 5$ fulfilled, but so also are those for all higher order modes. Thus the coexistence of several modes appears possible. Actually this does happen, and has a deleterious effect which has to be minimised; (see Sec.). However, the deformation of the space charge is greater the smaller the value of n so that, although other modes might well be present, the mode $n = 5$ would predominate in the case indicated.

To operate the magnetron in the $n = 4$ mode with fixed B would involve raising V_a so that the working point was (say) T. It is possible to show, however, that for efficient operation V_a should be as near as possible to the energy supply potential. In any mode for which the energy transfer potential considerably exceeds the energy supply potential, as with the $n = 4$ mode here shown, working would be inefficient. It would be better in the case shown to work in the mode $n = 5$ or $n = 6$.

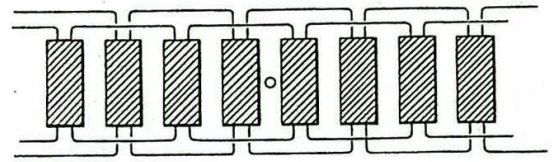
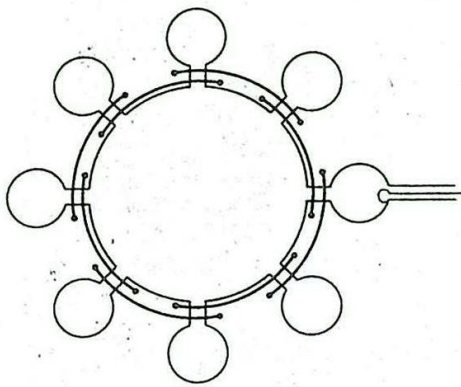
Consider V_a raised quickly, but not instantaneously, from zero to its working value. Then V_a rapidly passes through the conditions for oscillation in the modes given by $n = 10, 9, 8, 7$ etc., operation in these modes becoming more inefficient, and in some case ceasing altogether as V_a rises. These different modes will, in general, cause the magnetron to operate at different frequencies, so that operation at the desired frequency may not take place until some short time after the pulse has been initially applied. The ensuing delay between the application of the modulating pulse and the transmission of the RF pulse at the correct frequency might lead to Range errors if not allowed for in the timing system.

The coexistence of several modes, operating at slightly different frequencies is undesirable, and is minimised by Strapping.

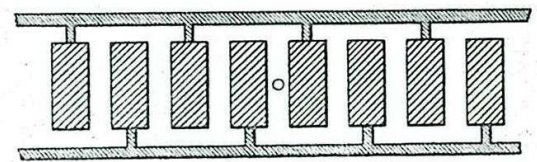
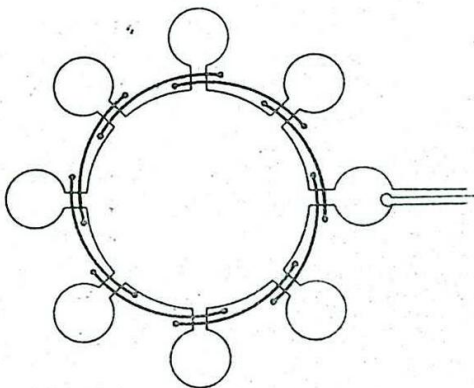
35. Strapping

The purpose of strapping is to minimise the coexisting modes, both in number and in magnitude of oscillation. It is done by connecting short conductors between pairs of segments, the segments so connected being separated by a third. In a general way it is easily seen that joining segments in this manner will tend to favour some modes rather than others. For example, if an 8-segment magnetron be taken and alternate segments joined as in Fig. 440, the π -mode ($n = 4$) is favoured, since this implies a phase difference between the strapped segments of 2π . For the mode $n = 5$ the phase difference would be $\frac{5\pi}{2}$, a condition impossible if the segments are short-circuited.

Strapping is, however, not as simple as is here implied, since the straps cannot be regarded as simple short-circuits. For each strap, being a conductor, has inductance, and since it passes over the separating segment, capacitance exists between it and this segment. Thus the presence of the strap alters the frequency at which a particular mode operates. In the π -mode the strapped segments are at the same potential, and hence the strap carries negligible current; so its inductive effect can be ignored. Its capacitive effect cannot; hence the frequency corresponding to the π -mode is lower when the magnetron is strapped than when it is not, by an amount largely determined by the strap position relative to the block. If the mode $n = 5$ arose there would be current through the strap; so its inductive effect would be important, as would its capacitive effect also. Hence although the general effect of strapping is to change the frequencies at which modes arise, such changes will usually be

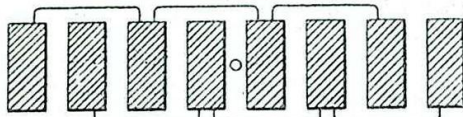


DOUBLE RING STRAPPING
(a) (SINGLE RING STRAPPING OMTS THE OUTER STRAPS AT ONE
END AND THE INNER STRAPS AT THE OTHER END)

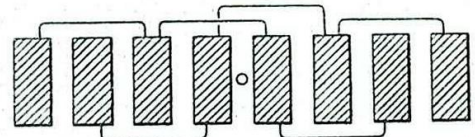


(c) STRIP STRAPPING.

(b) ECHELON STRAPPING
BOTH ENDS.



(d) INCOMPLETE RING STRAPPING.



(e) Y - B STRAPPING

Fig. 440 - Strapping systems.

different for the several modes. Thus the different modes are separated in frequency by strapping; and the conditions in the magnetron system as a whole, and in particular the selectivity of the load, may then be adjusted so that matching is achieved at the frequency of the required mode; (normally of the lowest order possible, and most commonly the π -mode).

The suppression (or minimisation) of all but the wanted mode markedly increases the efficiency. For example, it has been explained that, if V_a is such that the operating point is almost on the $n = 5$ energy supply potential line of Fig. 439, any $n = 6$ (or higher mode) output must be inefficiently produced. Hence its contribution to the total magnetron output must reduce the overall efficiency. Since strapping, by diminishing the unwanted modes and separating

them in frequency, tends to make the output consist almost entirely of the required efficient modes, its introduction results in a very considerable increase in the overall efficiency of this type of magnetron.

Referring again to Fig. 439, it has been stated that working in the $n = 4$ mode (at point T) would be less efficient than working in the $n = 5$ mode at point P. Whilst this is true for the unstrapped magnetron, it is not true for the strapped magnetron, to which Fig. 439 would not apply. This is because the value of the energy supply potential for a given B and n depends upon frequency, and hence changes with strapping.

Some common forms of strapping in practical use are shown in Fig. 440.

Strapping also provides a means of pre-tuning, i.e. setting up the magnetron on a spot frequency before it is sealed off. The lid is kept off and CW from a klystron oscillator at the desired frequency is injected into the magnetron through the coupling loop. The straps are then bent until the magnetron resonates at the required frequency, this condition being determined by the use of a resonance indicator in the coaxial line from the klystron to the magnetron. (Allowance has to be made for the difference in frequency occasioned by the presence of the lid and the space charge, in the operating magnetron; but empirical rules enable this to be done with the required accuracy).

The omission of one or more straps from an otherwise symmetrical system, or introducing a break in a strap (or straps) can have a marked effect on the response of a magnetron at some particular mode or modes. Strap omission is sometimes made necessary by such practical factors as the position of cathode leads etc.

36. Effect of Anode Length, and of End-Plate Distance

The length of the anode and the distance between it and the end plates influence the frequency of the various modes; Figs. 441 and 442 indicate the extent of this effect. The fact that the anode-to-end-plate distance affects the frequency of operation is made use of in tunable magnetrons, one form of which has a thin lid the centre of which may be depressed and the frequency thus changed. In this way a change of about 150 Mc/s in 3000 Mc/s can be achieved. Another form of tunable magnetron has metal rods which can move in and out of the anode block, thus altering both its effective length and the anode-to-end-plate separation.

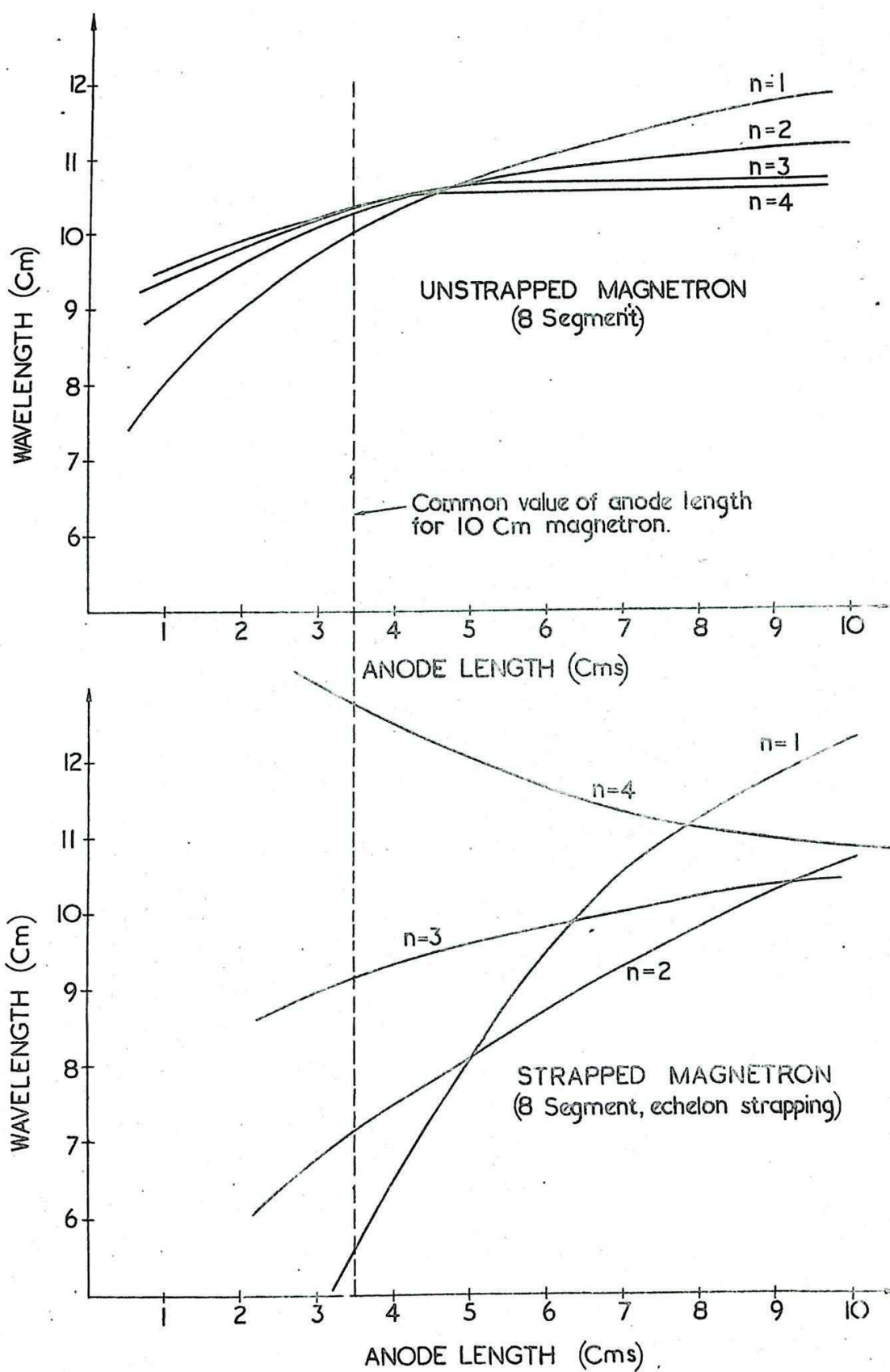


Fig. 441 - Wavelength v. anode length for strapped and unstrapped magnetrons.

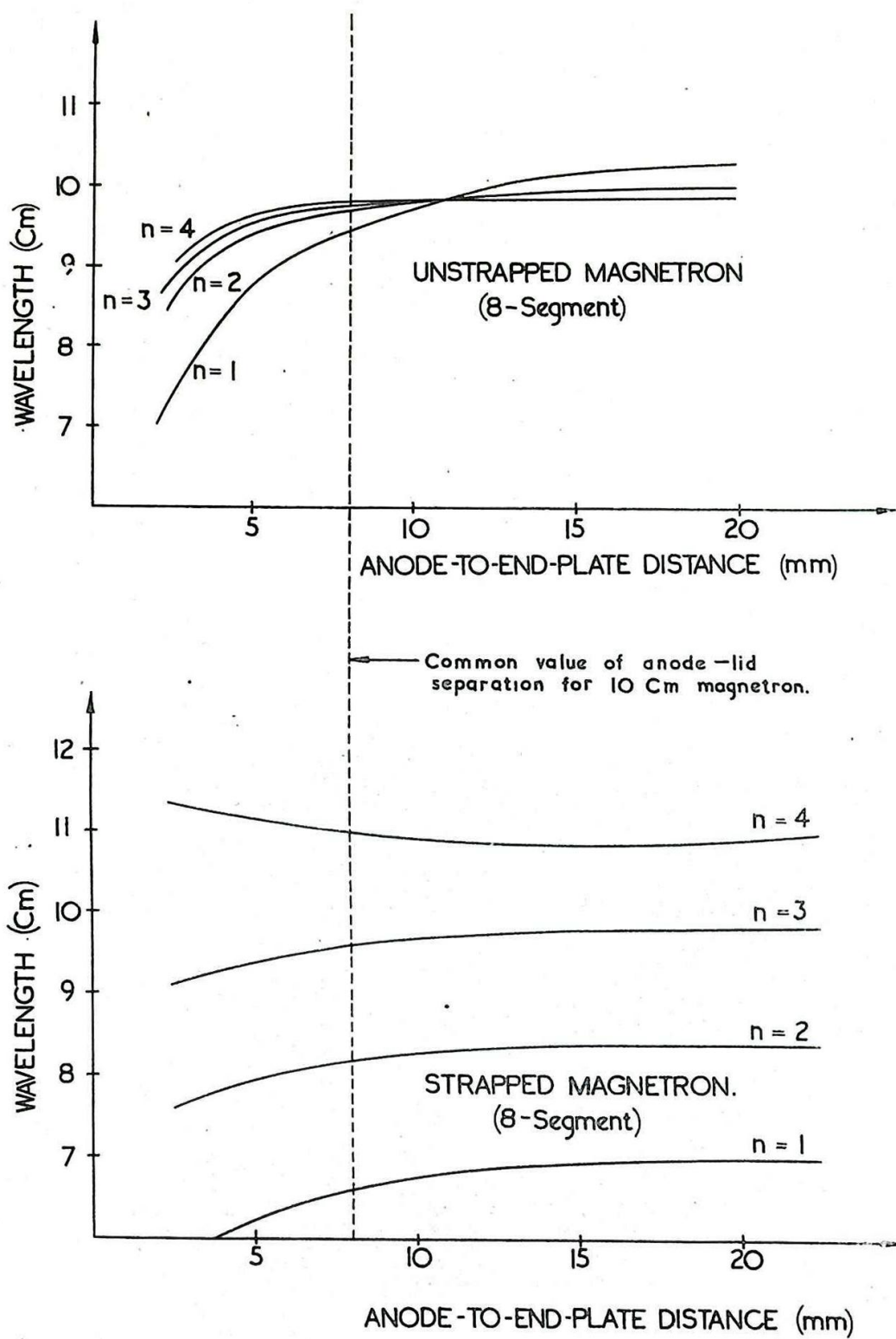
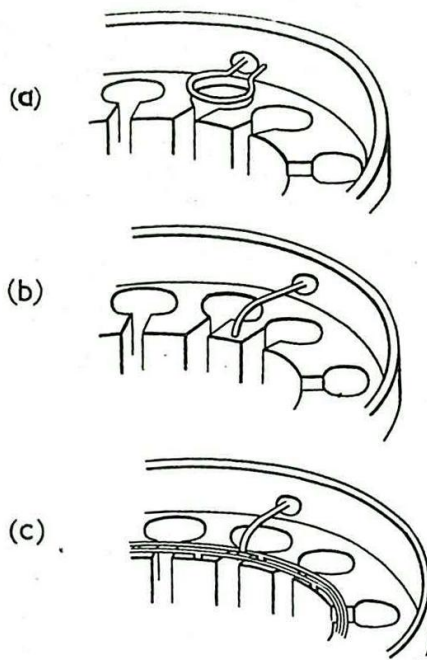


Fig. 442 - Wavelength v. anode-to-end plate distance for strapped and unstrapped magnetrons.

37. Output Couplings

The output is commonly taken from a magnetron by means of a coupling loop, made by bending round the inner of the coaxial output line and soldering it on to the outer (see Fig. 422). The output can then be adjusted by turning the loop, thus varying the magnetic flux linkage.

At 3 cm. and shorter wavelengths it is not practicable to make loops small enough to be located within the cavity in this manner, and the methods indicated in Fig. 443 are used. (a) shows a loop inductively coupled to the cavity; in (b) the output conductor is attached to the upper surface of one of the segments of the anode block; and in (c) to one of the ring straps commonly used at these wavelengths.



3 CM. OUTPUT COUPLINGS

Fig. 443 - 3 cm. output coupling.

38. Cathode Features

The turbulence of the electron cloud in an operating magnetron is vigorous enough to force a proportion of electrons back to the cathode with increased energy. There is consequently additional heat dissipated at the cathode and a possibility of cathode disintegration. The energy dissipated in this way is sometimes so large that, after an initial warming-up period, it is possible to reduce the magnetron heater supply, or even switch it off entirely. The cathode must be robust to withstand the bombardment, and is commonly made by using a nickel mesh as a foundation and filling this with emissive oxide.

39. Pulse Operation of Magnetrons

In most radar applications magnetrons are required to produce high power RF pulses of short duration, this being normally achieved by applying to the cathode approximately rectangular negative-going pulses, of the required duration. This is preferable to applying positive going pulses to the anode, which it is usually convenient to earth. The pulses should be steep-fronted and have a rapid die-away, to minimise the effect of unwanted modes set up when V_a is changing. (See Sec. 34). They should also have a flat top, since changes in V_a during the oscillating period can seriously affect frequency, efficiency and output power.

40. Magnetron Characteristics and Rieke Diagrams

Two methods are in common use for presenting magnetron performance, namely, Magnetron Characteristics and Rieke Diagrams. Magnetron Characteristics are drawn in Cartesian Coordinates, the modulating voltage applied to the magnetron in pulsed operation being graphed against the value of the valve current I_a during the pulse. Contours of constant flux density B and contours of constant RF output power, are drawn. Sometimes efficiency contours are also included.

An example of a set of magnetron characteristics is given in Fig. 444. It is seen from this that to obtain high efficiency it is necessary to have a high value of applied voltage and magnetic flux density. For constant B there is an optimum value for V_a beyond which the efficiency begins to fall.

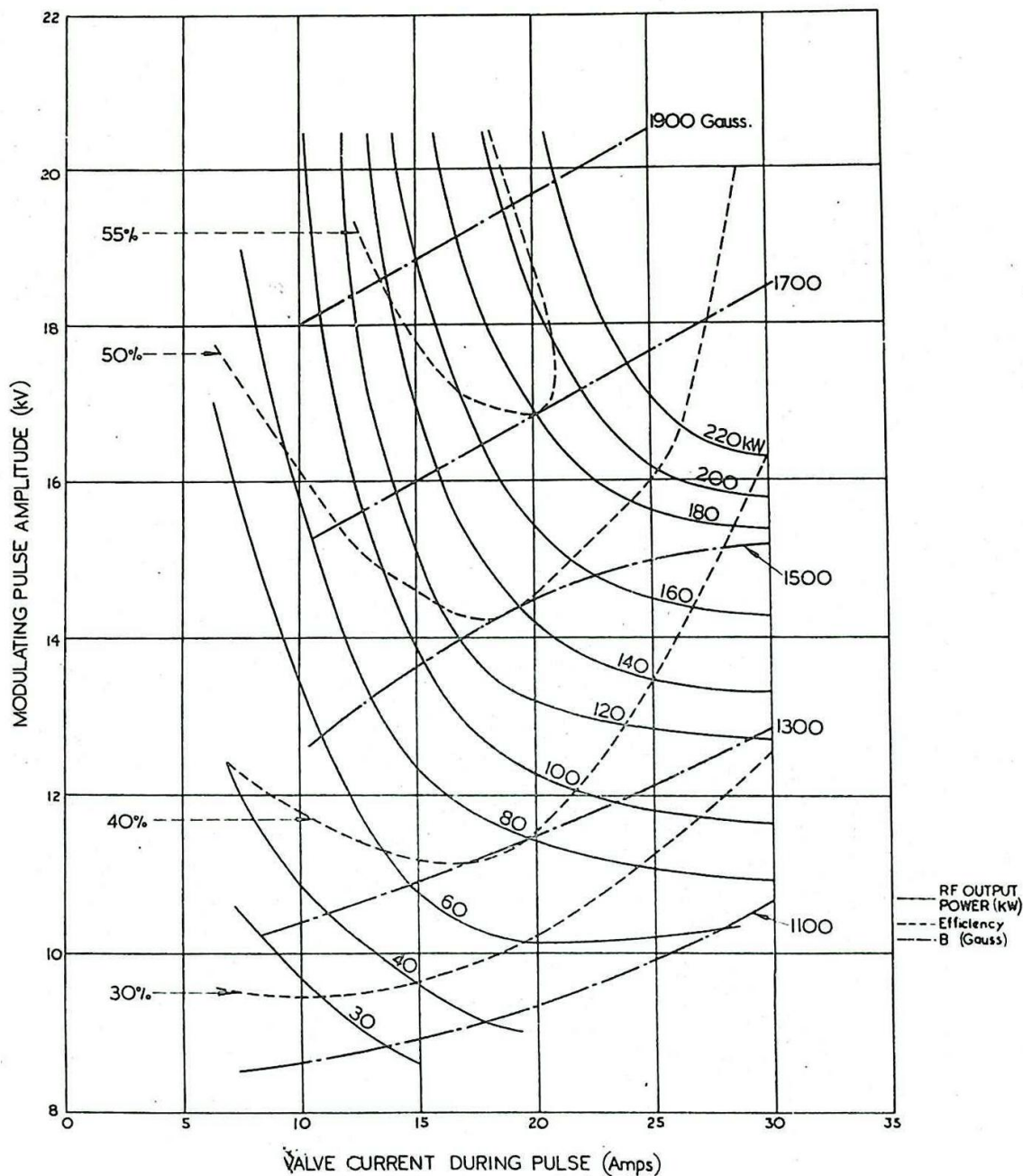


Fig. 444 - Magnetron characteristics.

Whereas Magnetron Characteristics indicate magnetron performance under conditions of varying V_a and B, the load being always adjusted for maximum output power, Rieke Diagrams show performance for fixed I_a (or V_a) and fixed B for variations of load impedance.

Rieke Diagrams

Consider the magnetic field strength and pulse amplitude to be constant. Then variation of load impedance over all possible values of resistance and reactance will cause changes in valve current, peak power output, efficiency, frequency of operation, and frequency stability. Corresponding changes occur if the current is fixed and the voltage varied. It would be possible to plot these variations as contours on Cartesian axes representing values of resistance and of positive and negative reactance; but it is more convenient to plot them on a Polar Circle Diagram as shown in Fig. 445. This has the advantage that the effect of a loaded output line or waveguide can be taken into account on the same diagram when performance is being deduced in a particular case. The normal circle diagram technique is then employed to transform the impedance to the output loop so that the corresponding performance can be read off.

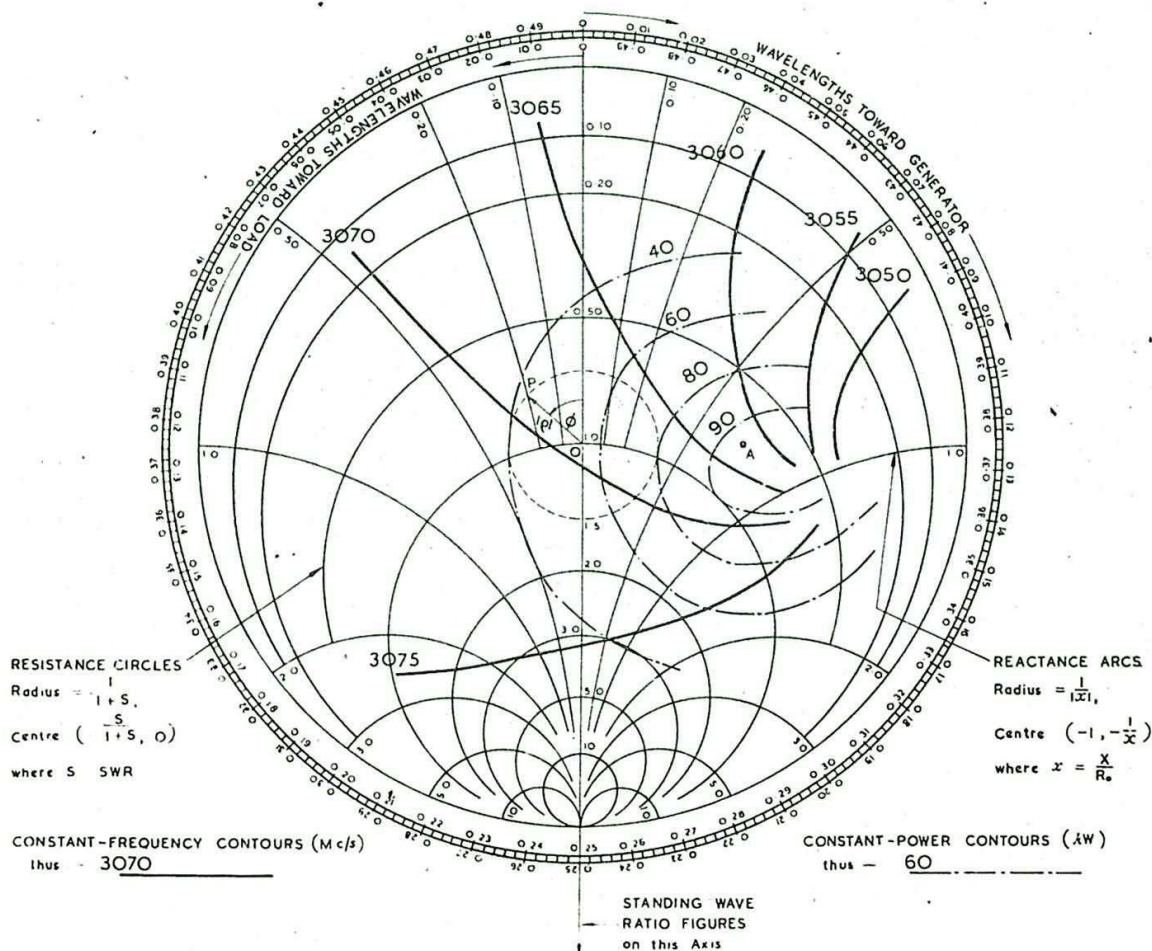


Fig. 445 - Rieke diagram.

In the Polar Circle Diagram, the complete circles represent circles of constant resistance and the arcs orthogonal to these circles, constant reactance. An important property of this circle diagram is that when the line is terminated in a load represented by the point P, the magnitude of the reflection coefficient ρ of the load is represented by OP, and the phase change on reflection is given by ϕ (see Fig. 444).

Hence circles centred at O are circles of constant ρ . These are also circles of constant standing wave ratio since reflection coefficient and standing wave ratio are interrelated ($\text{SWR} = \frac{1 + \rho}{1 - \rho}$). The vertical axis is scaled in terms of the standing wave ratio.

The following important points arise from the diagram. In the example taken, the region of highest power lies at A. Following the constant- ρ circle from A to the vertical axis reveals that the corresponding standing wave ratio is not unity. This implies that the line or waveguide is mismatched; and this is due to the fact that the optimum load for the magnetron at the output loop is different from the characteristic resistance of the line or waveguide. In order to match the magnetron output impedance to the line or waveguide, many modern magnetrons have a built-in matching device at the output loop. In the region of A the frequency stability is poor because the frequency contours converge, so that a small change in loading causes a comparatively large change in frequency. Hence operating the magnetron to obtain maximum output power makes good frequency stability impossible.

The operating conditions of the magnetron and the maximum efficiency at which it can be operated depend upon the characteristics of the magnetron and its load; i.e. the radar system of which the magnetron is part. The particular features of the radar system that are important are the modulating voltage and pulse shape, the type of receiver employed (e.g. whether there is Automatic Frequency Control or not), and the maximum standing wave ratio that may be set up in the feeder system due to joints, spacers, and variations in loading originating at the aerial array.

Suppose, for example, that the standing wave ratio on the line or guide is subject to variation between 1 and 1.5, due, perhaps, to a variable termination of the line caused by a faulty scanning system. This variation of the SWR of necessity causes fluctuations in the magnetron frequency. If automatic frequency control (see Sec. 49 - 53) is not employed at the receiver this may seriously affect the performance of the radar system. Suppose that a frequency change of 10 Mc/s is permissible. Then, since a SWR of 1.5 corresponds to $\rho = 0.2$, the magnetron can be operated under such conditions that the constant- ρ circle of radius 0.2 does not cover more than 10 Mc/s change in frequency contours.

41. Frequency Stability and Pulling Figure

Frequency changes in an operating magnetron can be traced to several causes. As has already been pointed out, variation in input conditions produces a change in frequency. As this variation also results in a change of anode current, the effect is called Current Pulling and is of the order of 0.1 to 1 Mc/s per ampere for 10 and 3 cm. magnetrons. Variation in anode block temperature also causes a frequency change because of the expansion of the block. The change is roughly proportional to the change in temperature and the coefficient of linear expansion of copper. Load variations are the

most important causes of frequency changes. The term Pulling Figure is defined as the difference between the highest and lowest frequencies obtained as the load is varied in any manner which does not cause the standing wave ratio to exceed its maximum permissible value. Normal values of pulling figure are of the order of 10 to 15 Mc/s.

If a long transmission line or waveguide is employed between the magnetron and the aerial system operation at more than one frequency may be possible, and the magnetron frequency may jump from one value to another more or less at random. This is commonly known as Frequency Splitting. It is avoided by keeping transmission systems as short as possible or by avoiding standing waves on long lines or guides.

Use of Rieke Diagrams to Illustrate Frequency Pulling and Frequency Splitting

In the example already quoted in Sec. 40, referring to Fig. 445, the standing wave ratio was assumed to be 1.5, so that $\rho = 0.2$. The frequency variation over the region covered by the circle $\rho = 0.2$ may be determined by interpolation between the frequency contours already drawn. The pulling figure may thus be shown to be about 6.5 Mc/s.

The Rieke diagram also illustrates how a short line is preferable to a long one. Suppose that the line (or guide) is of length l and that the standing wave ratio is 1.5. If the frequency is varied the electrical length $\frac{l}{\lambda}$ of the line changes, but to a

first approximation the standing wave ratio may be taken as constant. The working point on the Rieke diagram correspondingly traverses a constant- ρ circle as shown in Fig. 445. For any given position of the point on the circle the corresponding angle ϕ is, in fact, the phase angle between the direct and reflected waves on the line at the input terminals, and hence the relation between f and ϕ is given by

$$\phi = \frac{4\pi l}{\lambda} + \phi_r,$$

where ϕ_r is the phase change on reflection at the load. (In the case of a waveguide λ_g should be read for λ in the preceding expressions). We may thus deduce that

$$\phi = \frac{4\pi l f}{u} + \phi_r \dots\dots\dots(5),$$

where u is the velocity of propagation on the line (or the phase velocity in the guide). In Fig. 446, f is plotted against ϕ , giving the line XY.

Suppose now that the line is energised by the magnetron. Then there is the fundamental relation between f and ϕ already plotted in Fig. 446, obtained direct from the Rieke Diagram. The working conditions are therefore given by the intersections of the two curves. In the example taken there are three intersections, suggesting that there are three possible operating frequencies. However, only two are possible, because although the three points of intersection give three conditions of equilibrium, the one at A is unstable. For if some small change in the system were to occur, resulting in a small increase in f , then, from equation (5) ϕ is increased; as shown by the frequency pulling curve, this would result in a further increase in f . On the other hand, at B and C

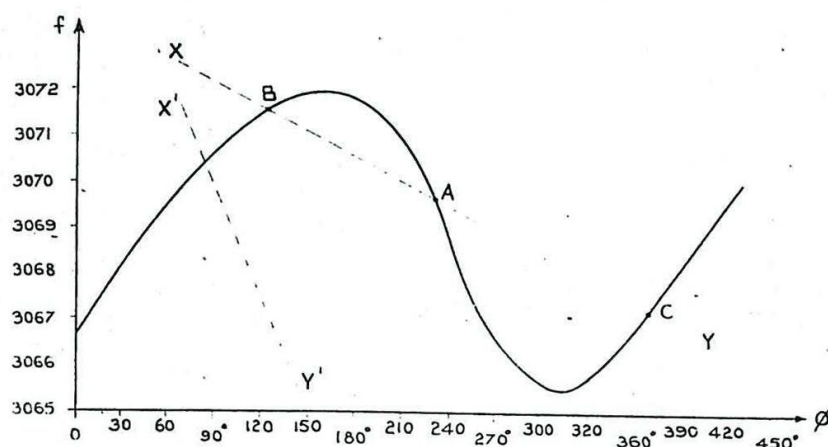


Fig. 446 - Frequency pulling curve (full line).

an increase in f causes a decrease in ϕ , which tends to pull the frequency back. Since these two frequencies are equally possible the magnetron may jump from one to the other, causing frequency splitting. If this is to be avoided the line XY must be sufficiently steep for the curves to intersect at one value of f only as indicated by $X'Y'$. The slope of the line XY is $\frac{u}{4\pi l}$; hence, to make XY steep enough to avoid frequency splitting, l must be sufficiently small.

For a certain critical length of line or guide the slope of XY will be equal to the maximum slope of the frequency pulling curve: and it is then possible, with one value of ϕ or r , for XY to be tangential to the frequency pulling curve at its steepest part. This leads to very considerable frequency instability, and is termed Broad Spectrum operation.

Further, a long line or guide is troublesome because the impedance as seen by the magnetron changes at time intervals corresponding to twice the transit time of the line or guide, due to reflections from the load back to the magnetron; these cause varying load impedance at the magnetron output loop and corresponding frequency changes. (The importance of this effect will depend upon the extent of the mismatch at the termination).

The longer the line, the more stringent must be the magnetron specification in regard to frequency stability; i.e. the smaller must be the pulling figure. An improvement in stability may be obtained even for long lines by putting a stub tuner close to the magnetron to adjust the region of operation so that it covers a fairly flat portion of the frequency characteristic.

42. Pre-plumbing of Magnetrons

Pre-plumbing is a process whereby magnetrons, complete with such devices as stubs, common T/R, and other switching apparatus, etc., can be so accurately manufactured that subsequent adjustment is unnecessary. This precludes any possibility of the introduction of instability due to maladjustment.

CONTROL OF NEGATIVE-GRID OSCILLATORS

43. Stabilisation of Oscillators

Oscillators may be divided into two classes, Master Oscillators and Power Oscillators. The former are usually required as frequency standards and are protected from variations in loading which would tend to upset the frequency stability. Power oscillators, on the other hand, must maintain an adequate frequency stability and a high degree of efficiency whilst undergoing various changes in loading conditions. Both types are used in radar. For various reasons, among which conservation of power, bulk and weight predominate, it is usual to employ power oscillators in transmitters rather than master oscillators followed by power amplifiers. At UHF there is no choice, since the problem of satisfactory amplification has not been solved. For calibrators, where accuracy is of primary importance, a frequency-stabilised oscillator is often used, followed by amplifying stages. Crystal-controlled oscillators are commonly employed, being superior to other types in this respect.

Oscillators should be protected from such fluctuations in supply voltage as produce variations in output frequency.

Control of the amplitude of the output voltage of an oscillator commonly takes the form of an automatic volume control circuit, controlling the bias of the valve. The use of automatic bias by grid leak and condenser, described in Chap. 7 Sec. 4, is a particular case. Frequency stabilisation may be provided by the use of a large-valued resistance in the feedback circuit. This resistance is in series with the slope resistance of the valve and its large value masks variations in R_a which might otherwise affect the oscillation frequency.

In some radio-frequency oscillators commonly used in super-heterodyne receivers and signal generators, the dual functions of master oscillator and power amplifier are performed by the same multi-electrode valve. A pentode, for example, may be employed, with control grid, screen and cathode acting as a triode oscillator, whilst the output is taken from the anode, screened from the oscillatory circuit by the suppressor grid. By this means variations in loading are substantially prevented from affecting the oscillation frequency. This type of circuit is usually referred to as an electron-coupled oscillator.

It is frequently found that a rise in mean anode potential affects the frequency in one direction, whereas a rise in screen potential has a reverse effect. When this occurs a suitable potentiometer arrangement for supplying screen and anode from the same supply may be chosen so that these two effects cancel each other, leaving the frequency of oscillations relatively independent of supply voltage.

In general it is possible to improve frequency stability by causing the circuit to oscillate at, or very close to, the resonant frequency of the tuned circuit employed. The effect of a high-Q circuit is largely nullified if it is used as a reactance. By inserting various compensating reactances in one or more of the electrode leads it is often possible to allow for interelectrode capacitances and other stray reactances so that the tuned circuit is used at the peak of its resonance curve. This is not practicable at very high frequencies when lead inductances introduce further complications.

44. Steady State Conditions

The performance of oscillators often depends fundamentally on the values of the supply voltages, and by suitable control of these an oscillator may be given various characteristics. To show how this is done it is first necessary to examine what factors determine the steady-state conditions of an oscillator.

Fig. 447(a) shows the schematic arrangement of the four oscillator components listed in Sec. 1. For purposes of description it is convenient to divide these into two parts, the "amplifier" and "network" sections respectively, as indicated at (b). Each of these parts may be considered separately, provided it is correctly terminated; i.e., when considering the behaviour of the amplifier section, it must be terminated by the input impedance of the network, and vice versa.

It is convenient to include the effect of grid current as extra damping in the network; this effect is of considerable importance in limiting the amplitude of oscillations in conventional oscillators.

It will be assumed that oscillatory voltages are sinusoidal. This is approximately true even for Class C operation, with non-sinusoidal currents, provided the tuned circuits are highly selective.

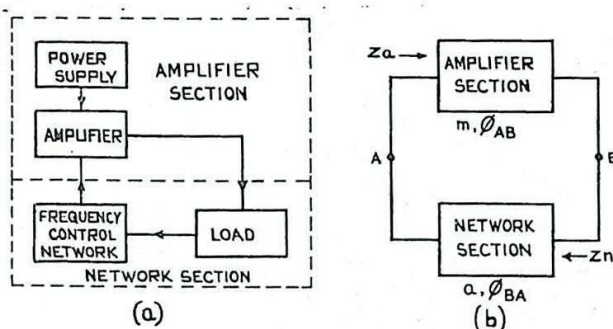


Fig. 447 - Schematic arrangement of fundamental oscillator components.

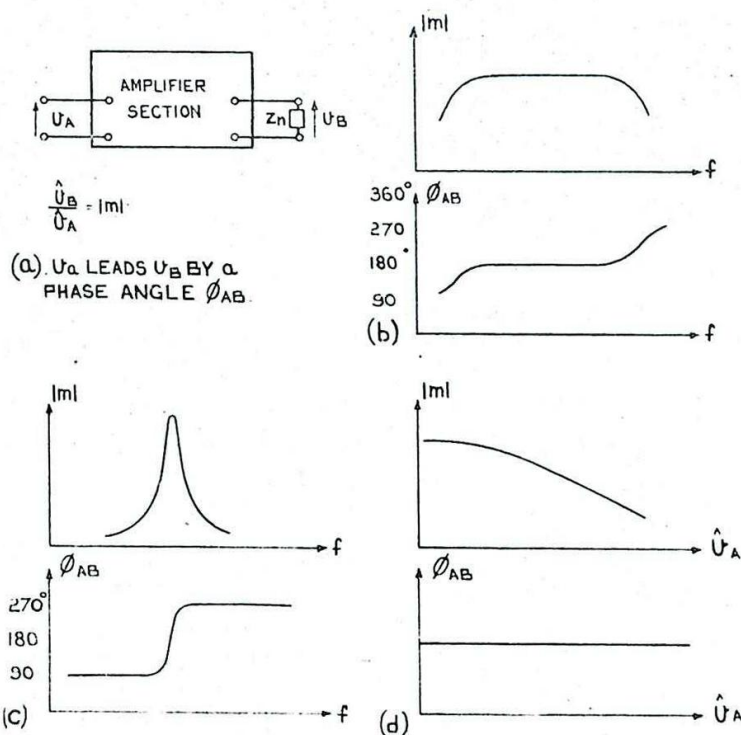


Fig. 448 - Amplifier section.

The behaviour of the amplifier section, terminated in z_n (Fig. 448(a)) may be described in terms of its amplification $|m| = \frac{\hat{v}_B}{\hat{v}_A}$ and phase shift ϕ_{AB} . Both of these vary with frequency. If z_n is resistive, as in a video amplifier, the normal variations of $|m|$ and ϕ_{AB} with frequency are as shown at (b), whilst if z_n is the impedance of a parallel tuned circuit, Fig. 448(c) shows typical variations.

As the amplitude of oscillation varies, the non-linearity of the characteristics causes changes in the value of $|m|$. To a first approximation ϕ_{AB} is not affected. These effects are illustrated at (d).

Similarly, the behaviour of the network section is illustrated in Fig. 449. If it is linear there will be no change in phase ϕ_{AB} or in attenuation $|a|$ as the amplitude is altered. $|a| = \frac{\hat{v}_B}{\hat{v}_A}$.

There will be considerable variation as the frequency changes, depending on the nature of the network. Fig. 449(a) shows a typical high-Q circuit which might form the network, and (b) and (c) show the attenuation and phase variations with frequency.

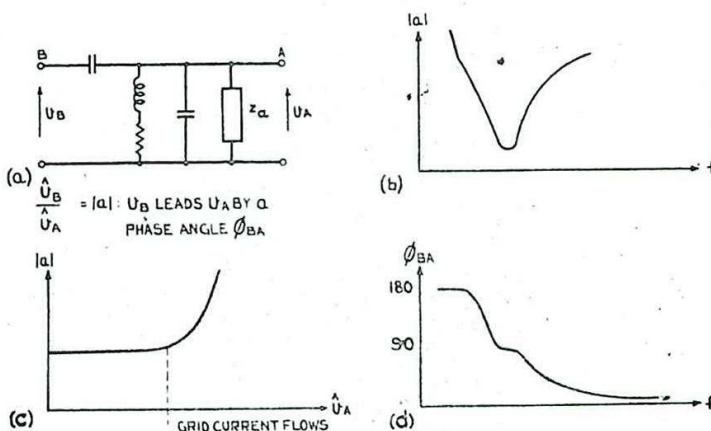


Fig. 449 - Network section (typical case).

When the effect of grid current damping is included in the attenuation $|a|$, this is not linear, but increases considerably when the amplitude of oscillations becomes large enough to raise the grid above cathode potential. This effect is illustrated at (d).

When the two sections, amplifier and network, are joined together to form the complete oscillator, steady oscillations can be maintained only if two conditions are fulfilled; the amplification of the amplifier must be just sufficient to compensate for the attenuation of the network; also the phase shift for the combined circuit must be a whole number of complete cycles; i.e. $|m| = |a|$, and $\phi_{AB} + \phi_{BA} = n \cdot 360^\circ$, where n is an integer.

Together these relations determine the frequency and amplitude at which oscillations may be maintained. In theory there may be several values of each which satisfy the steady state conditions, but in practical circuits one mode of oscillation usually predominates to the exclusion of others.

45. Effect of supply potentials on steady state conditions

In practical oscillators the frequency is seldom appreciably affected by variations in mean anode or grid potentials (save in the case of relaxation oscillators, not considered here.) But since such changes affect the amplifier gain they have considerable effect on the amplitude of oscillations. Maximum amplification is available under Class A working, but not necessarily maximum amplitude. This is because oscillations usually build up until the flow of grid current has increased the damping to the point at which attenuation is equal to amplification. An increase in bias in the first instance increases amplitude but decreases the values of both $|a|$ and $|m|$ in the steady state. If the bias is maintained at a steady level and the HT supply raised, amplitude is usually increased and also $|a|$ and $|m|$.

In most triode oscillators it is the effect of grid current rather than curvature of valve characteristics which predominates in determining the amplitude at which oscillations will be maintained.

With Class C bias, oscillations are not self-starting since initially the angle of flow of anode current (Chap. 7 Sec. 4) is zero, and therefore $|m| = 0$. When an input signal is applied causing anode current to flow, $|m|$ increases with amplitude as illustrated in Fig. 450(a), Curve III. The effect of increasing bias is to delay the point at which amplification commences and also to reduce the gain for a given amplitude. A similar result follows a reduction in HT potential, the reduced angle of flow causing a reduction in $|m|$. This similarity is illustrated at (b).

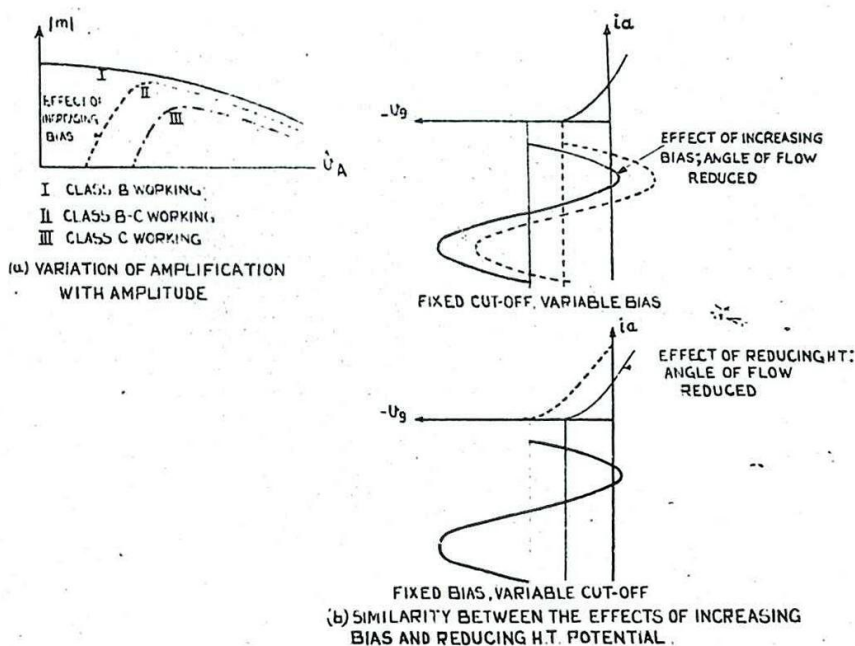


Fig. 450 - Effects of bias and HT potential on Class C operation.

The damping effect of grid current must be compared with the other damping factors of the network. Although the network may have a high Q , excessive grid-current damping will effectively reduce both the selectivity and the dynamic resistance of the network, assuming that it acts as a parallel tuned circuit.

In general, for a given HT supply, there is a unique value of bias necessary to maintain oscillations at a given amplitude. If a large amplitude is required this may be provided by an increase in bias. There is clearly a limit to this process, since as $|m|$ is reduced to a point at which it is less than the minimum attenuation $|a|$ oscillations cannot be self-maintained under any circumstances.

Similarly if the bias is held constant and the HT reduced, a point will be reached at which oscillations cease altogether, owing to the steady reduction of the angle of flow until $|m| < |a|$ for all values of \hat{V}_A .

46. Biassing circuits

As pointed out in Sec. 45, the effect on an oscillator of varying the bias is chiefly to change the amplitude of oscillations, and appreciable changes of frequency do not usually occur. If it is important that the amplitude should be constant, the steady supply potentials should not be allowed to vary, and the circuit should be designed as a master oscillator, so that changes in loading do not affect the operation. Under these circumstances, a constant bias can often be provided by a conventional cathode resistor with a by-pass condenser of sufficiently large capacitance, or by a suitable potentiometer arrangement.

If a valve oscillator is not otherwise protected from noticeable changes in loading, it should be provided with an automatic regulating device which increases the available power as the power demanded by the load increases. It will be shown that in common oscillator circuits such a requirement is not met by a cathode biasing circuit, and that for Class B or C working the reverse is true; viz. as the load increases due to a diminishing load resistance, which increases the damping of the circuit, the corresponding bias increases so that the available power is also diminished. In the circuit of Fig. 451(a) the valve current for Class B working is as shown at (b). Increased damping of the tuned circuit diminishes its dynamic resistance so that the amplitude of the anode alternating voltage is diminished; (c). This raises the anode voltage during the conducting period of each cycle, and thus increases the valve current; (b). This increases the bias. The same argument holds for Class C conditions.

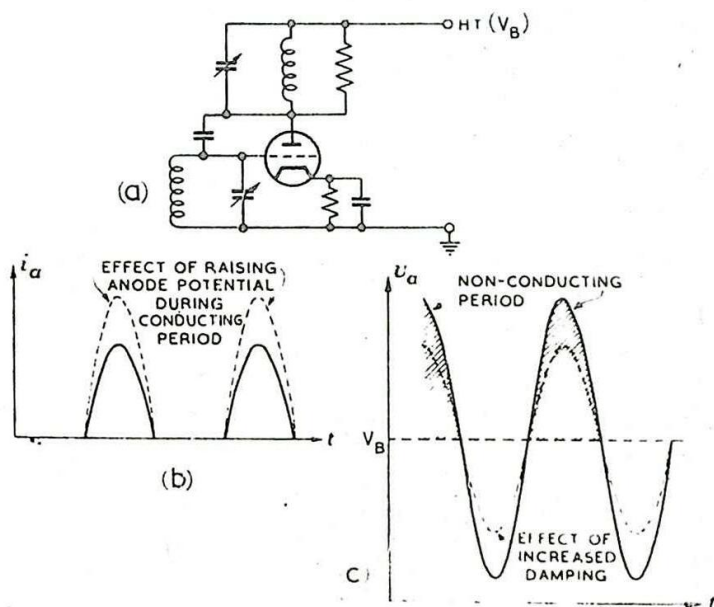


Fig. 451 - Non-self-compensatory effect of cathode bias.

The use of self-bias by grid leak and condenser provides automatic regulation approaching the form desired to compensate for variations in loading. In the circuit of Fig. 452(a) a decrease in amplitude of the alternating anode voltage decreases the amplitude of the grid voltage so that the condition for steady bias, viz, "charge accumulated on C during each cycle = charge which leaks away through R during each cycle" is no longer maintained, through the reduction in the angle of flow of grid current. More charge leaks away, diminishing the bias and increasing the angle of flow until a new stable condition is reached with a smaller bias. This increases the available power.

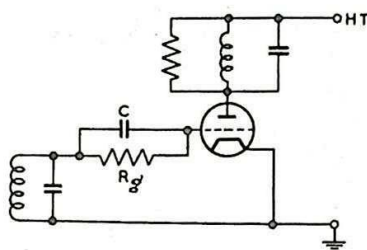


Fig. 452 - Biasing circuit using grid leak and condenser.

47. Self-quenching oscillators

The fundamental requirement for a pulse radar transmitter is the production of bursts of RF oscillations recurring at a suitable repetition frequency. This requirement can be met by suitable design of the power oscillator. Instead of maintaining continuous oscillations, an oscillator can be modified to allow oscillations to build up, reach a maximum, and decay, the valve remaining quiescent for a period, and the cycle then repeating automatically. This process is known as Self-Quenching, and when it arises from an auto-biasing arrangement in the grid circuits, the oscillator is very commonly known as a "Squegger", or Squegging oscillator. This term is also applied somewhat loosely to oscillators where the self-quenching circuit is in the anode or cathode lead.

In all cases the action arises from a gradual reduction in $|m|$ by an increase in the bias or a reduction of available HT potential, until $|m|$ becomes less than $|a|$. Under these temporary conditions oscillations cannot be maintained and usually cease altogether.

During the subsequent interval no valve current flows and the bias or HT potential returns to the value at which oscillations become possible, and the cycle repeats.

Squegging Oscillators

The actual oscillator circuit chosen to describe this behaviour is not important, and subsequent remarks apply equally to most types other than the common-cathode circuit drawn in Fig. 452.

The squegging action is illustrated in Fig. 453(a). Oscillations build up and the bias increases due to the accumulation of charge on C. Provided certain conditions are satisfied, which will be considered below, a point is reached at which the bias becomes greater than the value at which oscillations can be maintained, when the RF oscillations begin to be damped out. The charge on C leaks away through R_g and the bias is reduced until cut-off is reached. At this point $|m|$ is greater than $|a|$, and oscillations recommence. When grid current flows the bias increases and the cycle repeats.

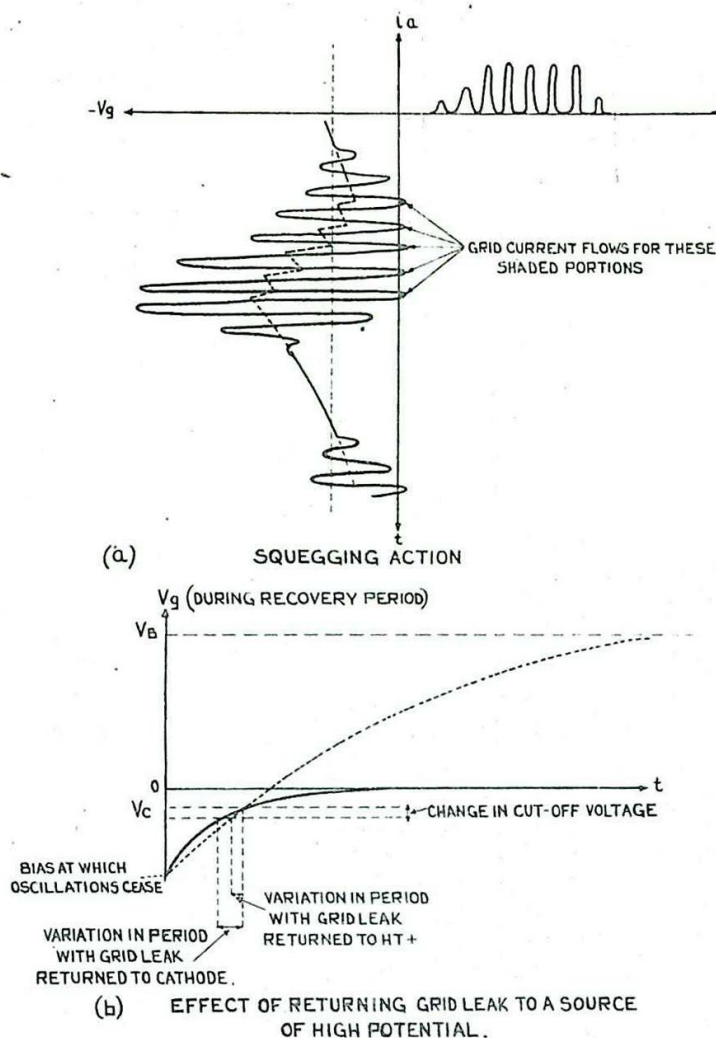


Fig. 453 - Effect of returning grid leak to a source of high potential.

The pulse width depends on the charging time-constant CR_i , where R_i is the mean DC resistance of the grid-cathode circuit when grid current flows. The recurrence frequency depends chiefly on CR_g , since it is usual for R_g to be much greater than R_i .

The circuit may be modified by returning the grid leak either to cathode or to HT potential. In the first case the action is the same except that the voltage developed across the grid leak is alternating instead of approximately constant. The connection of the resistor to a source of high potential, together with a suitable adjustment in time-constant so that the recurrence period is not altered, makes the onset of oscillations more definite and the recurrence frequency correspondingly more constant. This is illustrated at (b). A fluctuation in HT supply, for example, causing a variation in the cut-off voltage, would have much less effect on the periodicity in the modified circuit, as indicated in the diagram.

xxx Conditions under which Squegging Occurs

The mechanism of the cessation of oscillations may now be dealt with in greater detail. Suppose that the oscillations can be maintained at any chosen constant amplitude. The circuit is then equivalent to an amplifier being driven by a constant voltage generator. This generator will provide or absorb energy according

as $|m| < |a|$ or $|m| > |a|$ for this amplitude of signal voltage \hat{v}_i . For a given C-R network and constant input amplitude the bias developed across C will settle down to a steady value V_g depending on \hat{v}_i , C, R_i and R_g . Provided $CR_i \gg$ the duration of grid current flow for each cycle, and provided $CR_g \gg T$, the period of oscillations, the grid voltage will be approximately sinusoidal, as in Fig. 454, these being the conditions for non-distortion. The ratio $\frac{-V_g}{\hat{v}_i}$ then depends on the ratio $\frac{R_i}{R_g}$ only, to a close degree of approximation, and the relation between them is

$$\frac{-V_g}{\hat{v}_i} \approx \sqrt{1 - \left(\frac{\pi R_i}{2R_g}\right)^2} \quad \text{provided } R_i \ll R_g.$$

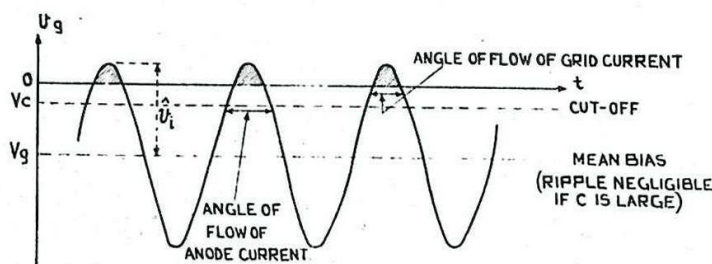


Fig. 454 - Steady state relation between \hat{v}_i , V_b and the angles of flow of grid and anode currents.

If we now plot the variation of $|m|$ and $|a|$ with amplitude \hat{v}_i , we obtain curves such as are shown in Fig. 455. The value of $|a|$ has been assumed constant, since the network is assumed linear, and the attenuation due to grid current is also linear under the assumed conditions.

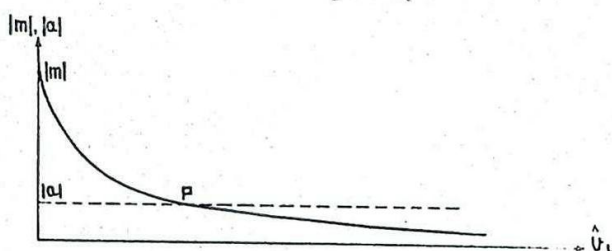


Fig. 455 - Variation of $|m|$ and $|a|$ with amplitude.

Provided the assumptions which have been made are justified, it follows that at the point P the controlling generator could be removed and oscillations would

be maintained at the corresponding amplitude, where $|m| = |a|$. That this does not happen in the squegging oscillator is due to the delay which occurs in the build-up of bias, since, before the steady state operating point P is reached, the bias will be less than that for which the curves of Fig. 455 are drawn. The effect of this lag is indicated in Fig. 456. Both $|m|$ and $|a|$ are increased, since the angles of flow of both anode and grid currents are larger. In other words, the amplitude will be momentarily stabilised at the value for the point Q instead of P, towards which it must tend for stability.

Two cases arise, shown at (a) and (b) respectively. In (a), $\delta |m| > \delta |a|$, and the point Q is to the right of P.

This means that the amplitude of oscillation is larger than the value at which $|m| = |a|$. Oscillations continue until the bias has increased beyond the required value, since this value must be exceeded before the amplitude begins to decrease. Once the bias has exceeded the critical value it remains too large for oscillations to be maintained, due to the long time-constant CR_g . Hence the oscillations die away.

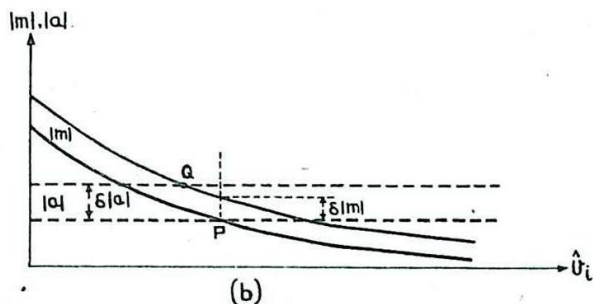
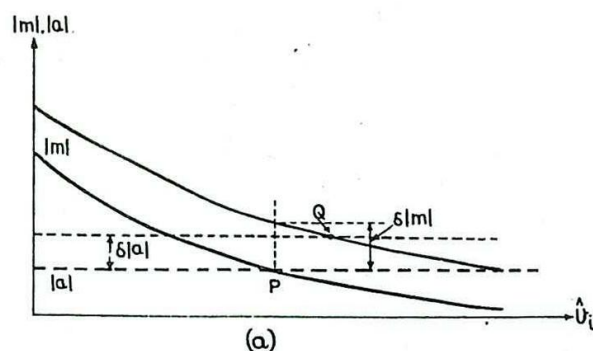


Fig. 456 - Effect of delay in build-up of bias.

In (b) $\delta |m| < \delta |a|$, so that the operating point Q is to the left of P. The amplitude of oscillations continues to build up with diminishing increases in bias, and the condition is stable, since each increase in bias makes $|a|$ increase with respect to $|m|$.

This is further illustrated in Fig. 457, where $|m|$ and $|a|$ are plotted against V_g for a fixed amplitude corresponding to the point P of Fig. 455. In the first figure (a) an increase in bias for a given amplitude leads to instability since $|a|$ becomes greater than $|m|$ and at the increased value of bias oscillations cannot be maintained. In (b), the increase in bias is accompanied by an increase in $|m|$ with respect to $|a|$, so that the amplitude is increased and the condition is stable.

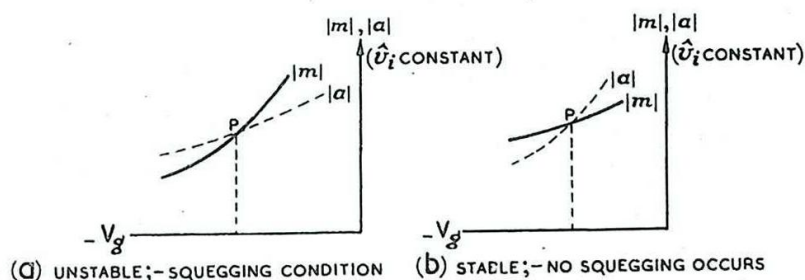


Fig. 457 - Squegging criteria.

The condition $\frac{\partial |m|}{\partial V_g} > \frac{\partial |a|}{\partial V_g}$ is the criterion for squegging

to occur, under the conditions which have been assumed (long time-constants, sinusoidal oscillations, and $R_g \gg R_i$). This criterion depends on the relative values of V_1 , R_i , R_g , the cut-off voltage

V_c and the dynamic resistance of the tuned circuit. For a given \hat{v}_i and V_c , and provided the dynamic resistance is not too large, squegging will occur if $\frac{R_g}{R_i}$ is sufficiently large.

Another method of making squegging more likely is to increase the damping of the tuned circuit. This causes the point P to occur at a steeper portion of the $|m| - \hat{v}_i$ curve (Fig. 456) and increases the likelihood of the point Q being to the right of P.

With triodes commonly used in low power radar oscillators employing high-Q circuits and Class C biasing, a grid leak of the order of 20 k Ω to 200 k Ω is usually sufficient to cause squegging.

Anode Self-Quenching Oscillators

In this form of self-quenching circuit the bias is held constant, but an effect similar to that of squegging is obtained by an automatic reduction in available HT potential. This reduces the angle of flow of grid current by shortening the grid base (Fig. 458(a)). A typical circuit is shown at (b). As in the grid self-quenching circuit, it is the value of the resistor R which is usually critical. If this is greater than a certain value quenching will ensue whatever the value of C. Pulse width is principally determined by the product of C and the mean DC resistance of the valve when conducting, whilst recurrence frequency depends chiefly on CR. The shape of the anode pulse is shown at (c).

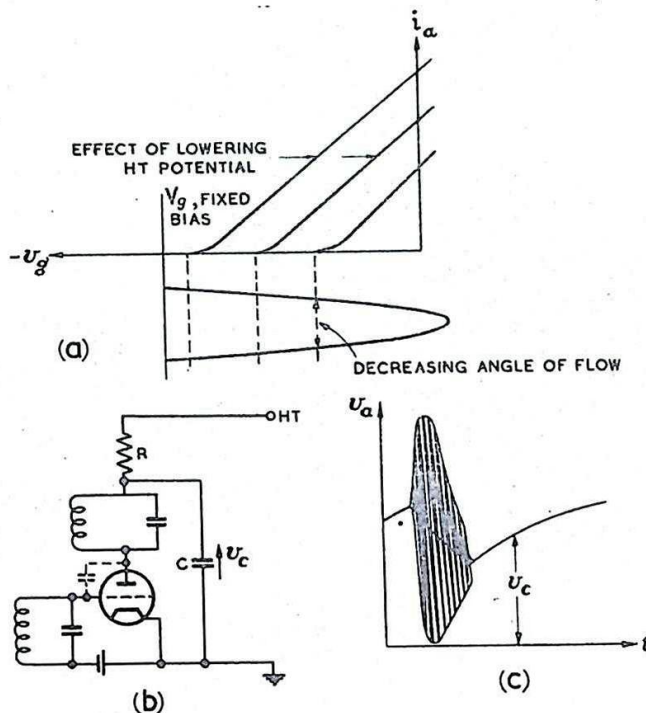


Fig. 458 - Anode self-quenching oscillator.

Anode self-quenching is normally employed as a protective device, rather than for regulating pulse width or recurrence frequency. Used in conjunction with grid self-quenching, the latter constituting the principal control, the anode network prevents damage to the valve should the grid circuit fail to prevent continuous oscillations.

48. The Super-Regenerative Receiver

Although the Super-Regenerative Receiver is more than an oscillator it is dealt with here since its operation depends on the build-up of oscillation in a modulated or quenched oscillator. The schematic arrangement of a separately-quenched circuit is shown in Fig. 459. (Self-quenched

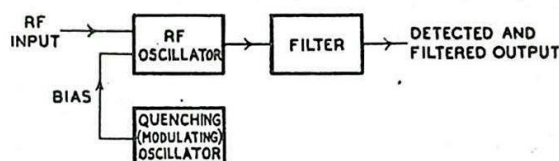


Fig. 459 - Schematic arrangement of super-regenerative receiver

receivers are sometimes used in communications sets, but are not commonly used in radar and are not described here). Since the schematic arrangement resembles that of a modulated transmitter it is as well to indicate the essential differences between the two circuits. It is generally assumed that in a pulse-modulated transmitter the build-up and die-away intervals for each pulse are very much shorter than the pulse duration. When the oscillator is used in a super-regenerative receiver, on the other hand, feedback conditions are so adjusted that in the absence of an applied signal the oscillations build up gradually and reach a maximum amplitude shortly before the end of the applied pulse; (Fig. 460). The oscillator is then quenched by the trailing edge of the modulation pulse and the oscillations die away rapidly.

The main difference between this action and that of the grid-modulated transmitter (Chap. 14) (or of the self-quenched oscillator described in Sec. 4.7) is that in the latter case the grid voltage rises sharply through cut-off so that the flow of valve current is sudden, initiating the main oscillation. In the case of the super-regenerative receiver, the applied pulse is less abrupt (being commonly sinusoidal) and the feedback less vigorous, so that the random noise fluctuations in the input circuit, superimposed on the applied pulse, are largely effective in determining the initial grid voltages during each oscillation interval. When a signal of the correct frequency is applied to the input circuit the mean anode current is increased and a filter circuit in parallel with the anode load enables the increase in current to be detected as a negative-going pulse.

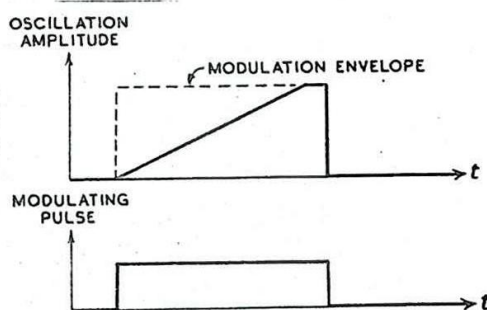


Fig. 460 - Build-up oscillations.

The effect of the super-regenerative circuit is that a given change of mean anode current can be produced by a very much smaller RF input signal than would be required, for example, by a reaction amplifier (Sec. 3, Fig. 385) used as an anode bend detector. Because of this a single valve stage used as a combined amplifier and detector in a super-regenerative circuit may provide an adequate output when the input signals are of the order of microvolts, whereas several hundred times this magnitude of signal would be required with a reaction amplifier-detector circuit.

The manner in which the much greater gain is obtained in the super-regenerative circuit may be explained with reference to the transformer-coupled circuit of Fig. 461. Suppose that the current and voltages in this circuit are zero when an EMF $v_i = \hat{v}_i \epsilon^{j\omega t}$, is injected into the tuned circuit by the coupling loop. The equation giving the output voltage v_o may be shown to be

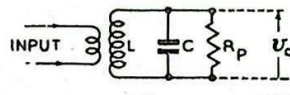


Fig. 461 - Transformer-coupled tuned circuit.

$$\frac{d^2 v_o}{dt^2} + \frac{1}{CR_p} \frac{dv_o}{dt} + \frac{v_o}{CL} = \frac{v_i}{CL} \dots\dots\dots(1).$$

The steady-state solution of this equation is the ordinary "AC solution", with $\hat{v}_o = Q\hat{v}_i$, where $Q = R_p \sqrt{\frac{C}{L}}$ (assumed large).

The manner in which the amplitude of the output voltage tends towards the steady state is illustrated in Fig. 462; the equation is:

$$\hat{v}_o = Q\hat{v}_i (1 - \epsilon^{-\frac{t}{2CR_p}}) \dots\dots\dots(2).$$

Suppose now that the resistance R_p is replaced by the equivalent of a negative resistance, $-R_n$.

The solution of the corresponding equation:-

$$\frac{d^2 v_o}{dt^2} - \frac{1}{CR_n} \frac{dv_o}{dt} + \frac{v_o}{CL} = \frac{v_i}{CL} \dots\dots\dots(3),$$

which is the same as (1) with $-R_n$ in place of R_p , is similar to that of the former equation, the amplitude growing according to the formula:

$$\hat{v}_o = Q\hat{v}_i (\epsilon^{\frac{t}{2CR_n}} - 1); \dots\dots\dots(4)$$

(in this case $Q = R_n \sqrt{\frac{C}{L}}$).

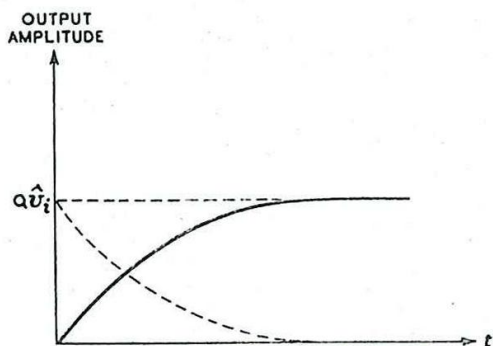


Fig. 462 - Build up of output voltage in transformer-coupled (non-oscillating) circuit.

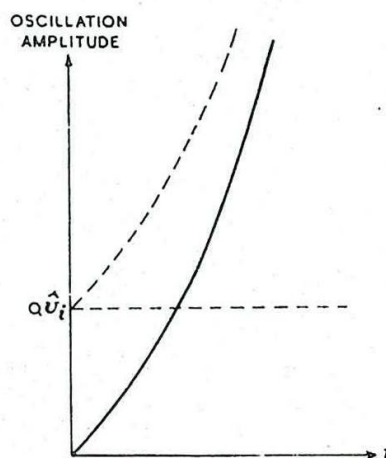


Fig. 463 - Build-up of oscillations in regenerative oscillating circuit.

This is illustrated in Fig. 463. Comparison of this figure with Fig. 462 shows that the former curve is the difference between the exponentially decreasing curve $Q\hat{v}_i e^{-\frac{t}{2CR_n}}$ and $Q\hat{v}_i$, whereas the latter curve is the difference between $Q\hat{v}_i$ and the exponentially increasing curve $Q\hat{v}_i e^{\frac{t}{2CR_n}}$.

If the amplitude of the applied signal is increased by an amount $\Delta\hat{v}_i$, the amplitudes of the curves of Figs. 462 and 463 are correspondingly increased. In the case of the regenerative (reaction) amplifier it is this increase in amplitude which provides the change in output current. In the super-regenerative circuit it is the time taken for the voltage v_o to reach a given amplitude which matters. When the oscillations have built up to a certain amplitude V_o , the amplitude remains approximately constant at this value, due to the inherent non-linearity and limiting action of the valve characteristics, until the end of the modulation pulse. This cuts off the valve current, in effect replacing the resultant negative resistance by the natural damping resistance of the circuit, so that the oscillations are rapidly quenched. The whole process repeats with the next modulating pulse, by which time the RF input voltage will have changed in amplitude.

Consider the action during a single modulation pulse. The equation relating the time T for the oscillations to build up with V_o , the maximum amplitude, is

$$V_o = Q\hat{v}_i \left(e^{\frac{T}{2CR_n}} - 1 \right) \dots\dots\dots(5).$$

Since we are concerned with values of T such that $e^{\frac{T}{2CR_n}} \gg 1$, for RF voltages of small amplitude \hat{v}_i this equation reduces to:

$$V_o \doteq Q\hat{v}_i e^{\frac{T}{2CR_n}} \dots\dots\dots(6).$$

If \hat{v}_c is the amplitude of the carrier wave modulated by $(1 + A \cos \omega_m t)$, we have

$$\hat{v}_i = \hat{v}_c (1 + A \cos \omega_m t);$$

so that

$$\frac{T}{2CR_n} \doteq \log_e V_o - \log_e Q - \log_e \hat{v}_c - \log_e (1 + A \cos \omega_m t) \dots\dots\dots(8).$$

Since, for a given carrier level, the only quantity which varies in (8) is t , we have

$$\Delta T \doteq -2CR_n \Delta \left\{ \log_e (1 + A \cos \omega_m t) \right\} \dots\dots\dots(9),$$

and this is independent of \hat{v}_c .

This decrease is illustrated in Fig. 464. t_p is the

duration of the modulating pulse. The two curves are identical except for the horizontal displacement ΔT . Hence, to a first approximation the average increase in amplitude of V_o due to the increase in the amplitude-modulated input is

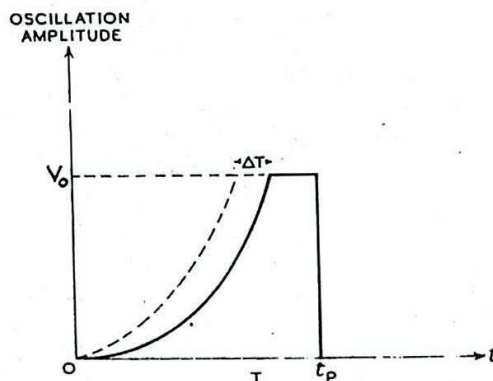


Fig. 464 - Decrease in time taken for oscillations to acquire a given amplitude (small signals).

$$\frac{\Delta T}{t_p} V_o ;$$

ie.

$$\frac{V_o}{t_p} \cdot 2CR_n \Delta \left\{ \log_e (1 + A \cos \omega_m t) \right\} \dots \dots \dots (10).$$

We shall compare this with the increase for a reaction amplifier, given by

$$Q\hat{V}_c \cdot \Delta \left\{ 1 + A \cos \omega_m t \right\} \dots \dots \dots (11).$$

Suppose the input changes from its maximum value $\hat{V}_c (1 + A)$ to its minimum value $\hat{V}_c (1 - A)$. For the reaction amplifier equation (11) gives

$$\Delta \hat{V}_c = Q\hat{V}_c \{ 1 + A - 1 - A \} = 2AQ\hat{V}_c .$$

For the super-regenerative circuit the change in mean amplitude for the pulse interval t_p is given by equation (10):-

$$\begin{aligned} \frac{V_o \Delta T}{t_p} &\doteq \frac{V_o 2CR_n}{t_p} \left\{ \log_e (1 + A) - \log_e (1 - A) \right\} \\ &\doteq \frac{2V_o CR_n}{t_p} \log_e \left\{ \frac{1 + A}{1 - A} \right\} . \end{aligned}$$

Provided A is small,

$$\log_e \left(\frac{1 + A}{1 - A} \right) \doteq 2A .$$

Hence, for a lightly modulated carrier, the ratio between the change in mean amplitude in the super-regenerative circuit to the corresponding change in mean amplitude in the reaction-amplifier circuit is given approximately by:-

$$\frac{2V_o CR_n 2A}{t_p} \div 2AQ\hat{v}_c = \frac{2V_o CR_n}{Q\hat{v}_c t_p}$$

Since $Q = \omega_o CR_p$, this may be written

$$\frac{2V_o CR_n}{\hat{v}_c CR_p \omega_o t_p} = \frac{2V_o R_n}{\omega_o t_p \hat{v}_c R_p}$$

Suppose $t_p = 1/\mu s$, $\omega_o = 2\pi 200 \cdot 10^6$ (corresponding to a frequency of 200 Mc/s); $V_o = 10$ volts and $\hat{v}_c = 10$ microvolts. Suppose also that $R_n = R_p$.

Then the ratio becomes

$$\frac{2 \cdot 10}{2\pi \cdot 200 \cdot 10^6 \cdot 10^{-6} \cdot 10 \cdot 10^{-6}} = \frac{10^6}{200\pi} \doteq 1690.$$

This ratio does not take into account the fact that the super-regenerative detector is operating for only a fraction of the quench period. However, even if the pulses fill only one-tenth of the reception time there is still an increase of more than 150:1 in the effective amplification provided by the super-regenerative circuit compared with the reaction amplifier.

The above simplified analysis has considered the action from the point of view of a sinusoidally modulated carrier - the communications aspect. In radar, where RF pulses are received, a somewhat different approach is simpler. Here we are concerned with the amplitude of the actual output pulses, and not with any modulation component of these; and the result found to be approximately true for a sinusoidally modulated carrier, namely, that the amplitude of the modulation component is independent of the carrier amplitude, no longer applies. There is still however, an inherent AGC action, independent of this result.

If the amplitude of the RF input is large, the approximation of equation (6) no longer holds, and the change in the rate of increase of oscillation amplitude following an increase in input amplitude is as shown in Fig. 465, rather than that shown in Fig. 464 for small signals. As the amplitude of the applied signal increases the increase in the mean amplitude of the output voltage does not increase proportionately, but is obviously limited. There is a

maximum possible increase in the area under the modulation envelope of Fig. 460, occurring in theory when the envelope is entirely filled with RF oscillations. The result is that the super-regenerative receiver can detect, without ever overloading, signals varying in amplitude from a few microvolts to several volts.

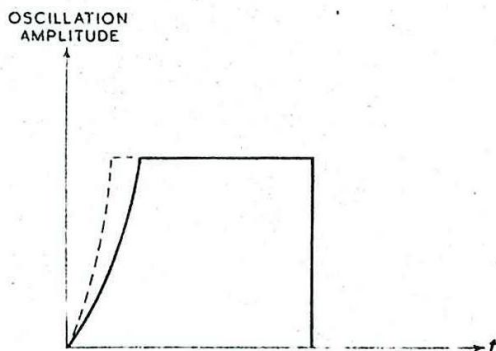


Fig. 465 - Decrease in time taken for oscillations to acquire a given amplitude (large signals).

When used for communications reception the region of variation of input amplitude over which distortion is not prohibitive is limited, and varies with the modulation depth. For radar purposes, where this type of circuit is used as a one-valve receiver for triggering radar beacons, distortion is not important. The output is of sufficient amplitude, even for very small input signals, to trigger a local transmitter, and because of the inherent AGC action no deleterious effects arise when much larger input signals are received. However, additional AGC circuits may be employed with advantage to limit the amplitude of the triggering pulse applied to the transmitter and to prevent valve saturation when very large signals are received.

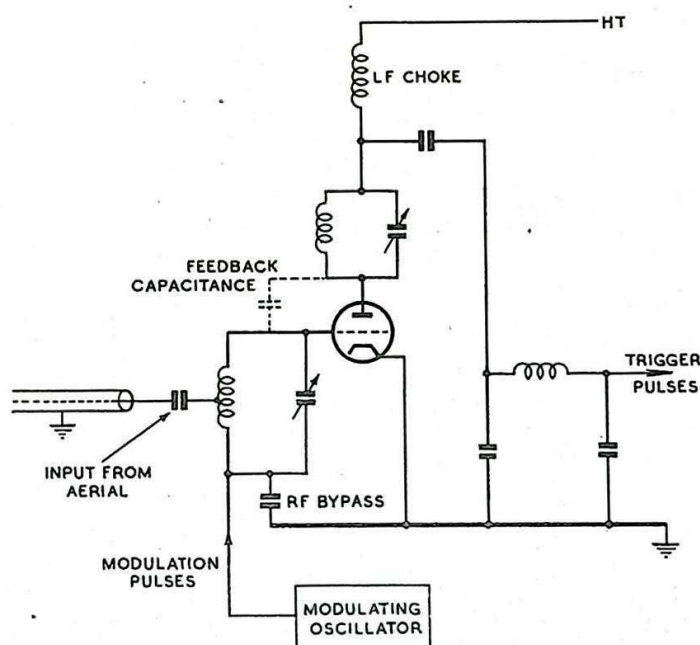


Fig. 466 = Typical radar super-regenerative circuit.

A typical arrangement of a super-regenerative receiver designed for RF pulse detection is shown in Fig. 466. The low-pass filter in the output side by-passes the fundamental and harmonic components of the quench frequency so that there is a constant steady voltage produced at the output terminals when no RF signal is applied to the grid. There is an inevitable noise fluctuation present. The width of the modulation pulse is made small and the quench frequency sufficiently high to ensure that several quench periods occur during a single input RF pulse.

When this arrives at the grid a negative-going pulse is produced at the output, slightly delayed by the filter circuit. This pulse is amplified and is used for triggering the local transmitter of a radar beacon.

AUTOMATIC FREQUENCY CONTROL

49. General

Like most automatic control systems, circuits which control the frequency of an oscillator are of the serve type (see Chap. 18). The frequency of oscillation is compared with some frequency standard in a suitable difference element, the output of which is used to adjust the oscillator so as to reduce the frequency

difference to zero.

The simplest requirement, namely to lock the oscillator frequency to that of an available transmitter or standard generator will not be dealt with here. In effect this converts the oscillator into an amplifier by increasing the coupling between the two generators. The more important practical requirement is to make automatic adjustments of the frequency of the local oscillator of a superheterodyne receiver so that the signal in the IF channel is kept in the middle of the preset IF band. This is necessary to prevent distortion and loss of signal strength. It is immaterial whether the changes in frequency are due to changes at the transmitter, which is more likely, or at the local oscillator. In either case, when such variations are detected, adjustments are made to the local oscillator to bring the IF signal back into the middle of the band.

The input quantity of the control system is therefore the setting of the IF tuned circuits, and the servo action is required to pull-in the IF signal by adjustments to the local oscillator. This may be done in the case of mechanically tuned oscillators by a servo motor turning the vanes of a variable condenser, altering the position of a tuning core or, in the case of a klystron, adjusting the tuning screws or diaphragm. Alternatively, if the frequency can be varied by altering one of the control voltages, such as the reflector potential of a Sutton tube, the motor can be dispensed with. In either case the chief problem in design is afforded by the difference element, usually called the Discriminator. This is required to convert a frequency difference into a voltage or current, indicative not only of the magnitude of the difference but also of the sense. Apart from the discriminator the remainder of the circuits involve nothing peculiar in servo construction.

50. The Discriminator

Two types of discriminator are employed, using amplitude and phase discrimination respectively. These are illustrated in Fig. 467. f_m is the mean intermediate frequency to which f_i , the IF signal frequency is to be aligned.

In both cases the reception of RF pulses within the bandwidth of the radio receiver develops corresponding pulses in the output of the discriminator which are either positive-going or negative-going according to whether the IF signal frequency is too high or too low. These pulses may either be rectified and applied as steady control voltages to the local oscillator or, as is more general, they may be incorporated with some automatic sweep circuit as described in Secs. 53 and 54.

51. Amplitude discrimination

In this method use is made of the variation in amplitude of the output from a high-Q circuit as the input frequency recedes from resonance. The circuit is tuned to $f_m \pm f_d$ where f_d is the maximum deviation allowed for. As f_i varies, so the output amplitude varies as indicated by the resonance curve (Fig. 467) and the detected output is thus indicative of the input frequency.

One method of incorporating this technique in a practical discriminator is shown in Fig. 468. Two high-Q rejector circuits, coupled to one of the IF stages, are tuned to $f_m + f_d$ and $f_m - f_d$ respectively. The output from each circuit is detected by the back-to-back diode circuits shown, and the voltage difference

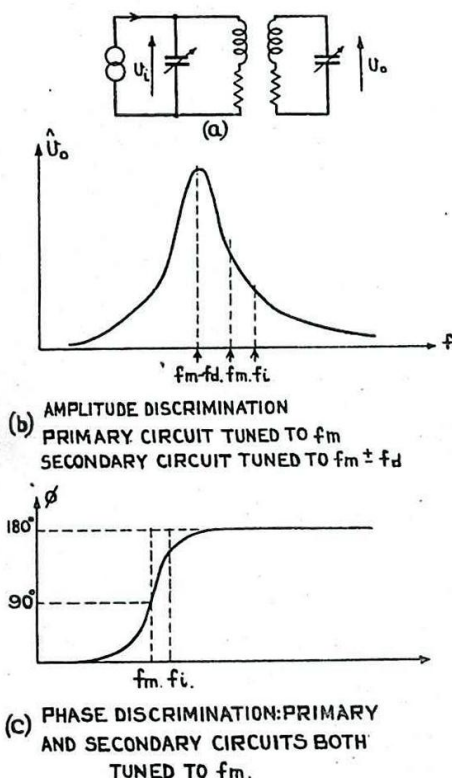


Fig. 467 - Fundamental principles of amplitude and phase discrimination.

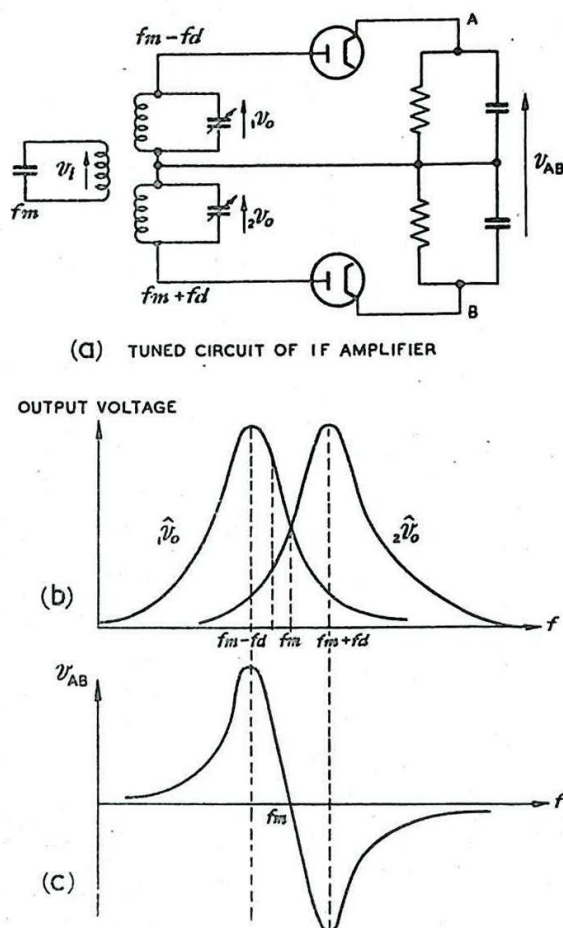


Fig. 468 - Amplitude discrimination circuit and response diagrams.

between A and B is indicative of the magnitude and sense of the error frequency $f_i - f_m$. The response diagrams for the two circuits are shown superimposed at (b). The detected output difference (c) is amplified if necessary and fed to the servo motor or other control device. This alters the frequency of the local oscillator so as to reduce the magnitude of the error.

52. Phase discrimination

As the title indicates, in this method it is the variation in the phase of the output of a tuned circuit as the input frequency varies which is used to provide the necessary frequency discrimination. It is necessary to compare this with a standard phase, normally the phase of the input. In the simple inductively coupled circuit shown in Fig. 467(a) the output and input voltages are in quadrature at resonance, the output phase being retarded as the frequency rises. If input and output signals are added together this change in phase is converted into a change in amplitude, as indicated by the vector diagrams of Fig. 469.

The resultant signal is detected and the magnitude of the rectified signal, varying with the input frequency, provides the discriminator output.

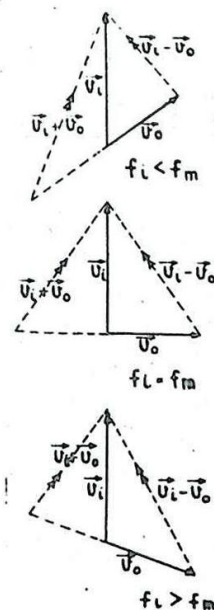


Fig. 469 - Phase discrimination: vector diagram.

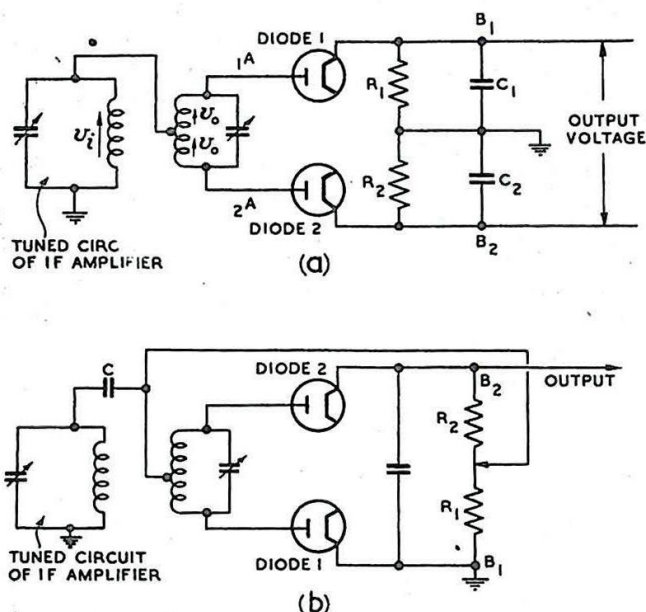


Fig. 470 - Phase discriminating circuits.

The simplified diagram of Fig. 470(a) shows how this method may be employed in a phase discriminator. v_i is the input voltage and v_o the voltage developed across each half of the secondary coil. As $1A$ swings positive the RF potential $v_i + v_o$ is developed across diode 1 and it conducts, charging C_1 to a peak value ($v_i + v_o$). Similarly diode 2 conducts as $2A$ swings positive so that C_2 charges to a peak value ($v_i - v_o$). R_1 and R_2 are the discharge resistors for the two condensers. The potential difference between B_1 and B_2 is $(v_i + v_o) - (v_i - v_o)$. In the circuit shown this is not usually in a convenient form for use as a control voltage, neither B_1 nor B_2 being earthed.

A practical form of this circuit is shown in Fig. 470(b). Here the coupling capacitance C takes the place of one of the rectifying condensers and the other is connected between the two cathodes. The direct connection from C to the junction of the resistors is necessary to discharge C through R_1 when diode 1 is not conducting. Also, if a variable potentiometer is substituted for the two resistors, this connection forms a convenient set-zero adjustment.

As in the simplified circuit, the output is developed across B_1B_2 , and since B_1 is earthed this is in a convenient form to act as a control voltage, either direct or through a DC amplifier.

53. Automatic Sweep Circuit

Either of the methods described is restricted because of the narrow bandwidth of the high-Q circuits used. The twin-circuit method of amplitude discrimination gives a wider band, but this is compensated for by the relative intricacy of setting up the circuit. Unless the oscillator frequency is very close to its required value, the IF signal is outside the discriminator bandwidth and the AFC circuit fails to "pull-in". The difficulty is usually overcome by incorporating an automatic frequency sweep circuit in the control system. While there is no signal in the RF receiver the local oscillator frequency is swept continuously through a wide band.

When a signal does appear, the AFC circuit comes into operation as the oscillator frequency reaches its correct setting, and thereafter the latter is automatically adjusted and the sweep practically eliminated.

A circuit incorporating a phase discriminator and a low frequency sweep generator is illustrated in Fig. 471. The output from the R-C filter circuit following the multivibrator is approximately sinusoidal. While there is no RF signal the anode voltage of the amplifier remains constant and the full sweep voltage is impressed on the control valve, which alters the Sutton tube frequency by varying its reflector potential.

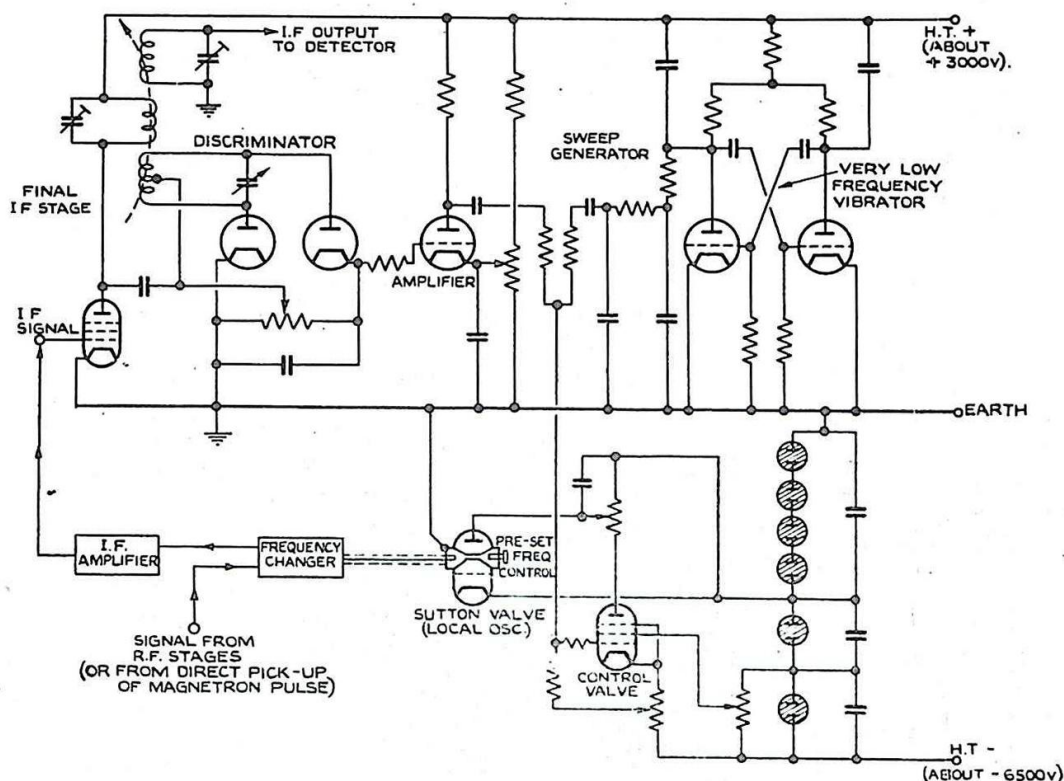


Fig. 471 - Circuit schematic for AFC of electrically tuned local oscillator.

When an RF signal appears the discriminator is brought into play and its amplified output nullifies the output from the R-C filter so that the control grid of the control valve is maintained at an approximately constant potential. If the discriminator is sufficiently sensitive and the gain of the amplifier sufficiently large, only a very small frequency deviation is necessary to produce at the amplifier output sufficient amplitude to counteract the sweep voltage.

This method of searching for and locking onto a received signal is applicable to other local oscillator circuits in which variation of the potential of one of the control electrodes provides a sufficient degree of variation in the oscillation frequency.

54. Practical AFC circuits

Many different AFC circuits are in common use, most of which incorporate the principles of the foregoing paragraphs, differing only in the manner in which these principles are applied. A great

variety of automatic sweep circuits are available, some of which are much more economical of material than the relatively straightforward circuit shown in Fig. 471.

It is common practice to apply the positive-going or negative-going pulses from the discriminator to the sweep generator so that it is held inoperative, automatically providing the correct output voltage for the local oscillator reflector. This usually necessitates suppressing either the positive-going or the negative-going pulses, so that adjustments to pull in the oscillator frequency are made for frequency drifts in one direction only. For example when the discriminator output is weak, or consists of negative-going pulses, the sweep generator may continue to sweep the oscillator throughout a wide frequency range. When the discriminator output changes to a series of positive-going pulses the sweep generator is suppressed and the oscillator frequency automatically pulls in. Should the transmitter frequency drift so that negative-going pulses reappear in the discriminator output the AFC circuit fails to pull in, and the sweep generator comes into operation again until the discriminator output changes polarity. This mode of operation is particularly useful where economy of valve stages is important.

CHAPTER 9

VALVES AS AMPLITUDE LIMITERS

1. GENERAL PRINCIPLES

In the normal operation of a pulse amplifier, described in Chap. 7, Secs. 9-13, care is taken that the valve works on the linear portion of its dynamic characteristic, otherwise amplitude distortion occurs. As an extreme case of this distortion, Amplitude Limitation arises when the amplifier is operated over one or both of the flattened portions of the dynamic characteristic (Fig.472).

Common applications of amplitude limitation are:-

- (i) The production of a rectangular output of voltage from a sinusoidal input.
- (ii) The elimination from the output of irregularities which occur at the extreme values of the input voltage.
- (iii) Discrimination between wanted and unwanted pulses.

Either diodes or amplifying valves may be used for amplitude limitation. In the case of the diode circuit of Fig.473

the maximum value of the output voltage is approximately zero, so that the positive portions of the input voltage are ineffective. By a suitable biasing arrangement this limiting action can be made to take place at any desired voltage level.

Two diodes are necessary if both upper and lower excursions of the output voltage variations are to be limited.

With a single amplifying valve both positive and negative excursions of the output voltage may be limited if the valve is operated over both of the flat regions of the dynamic characteristic. Alternatively, the grid-cathode portion of the valve may be used as a diode, and the positive part of the input limited when grid current flows by the inclusion of sufficient series resistance, as in Fig.473. This takes the place of limiting at the upper bend of the dynamic characteristic, and limits the negative-going portions of the output

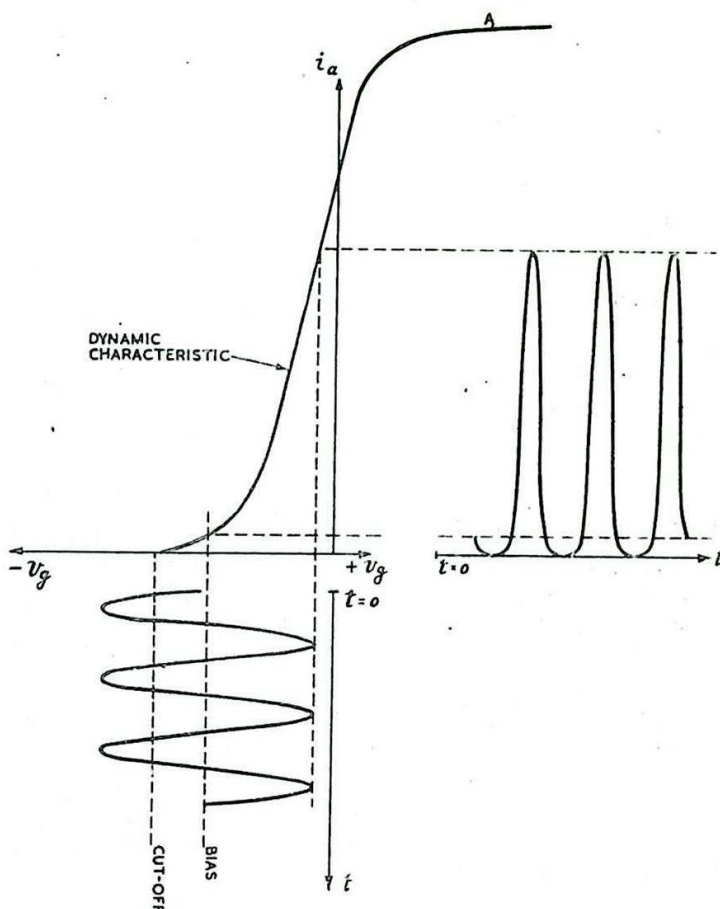


Fig.472 - Limiting at cut-off

voltage. In all cases the limiting action is independent of the nature of the input voltage variation; for simplicity, in discussing the action of the elementary circuits we shall assume that this is sinusoidal.

2. DIODE LIMITING CIRCUITS

A diode circuit is considered first, as it provides the simplest case of the use of a valve as an amplitude limiter. Suppose that a sinusoidal voltage is applied to an unbiased diode in series with a resistor R_s (Fig. 473). The diode conducts during the positive half-cycles of the input voltage and does not conduct during the negative half-cycles. If R_s is large compared with the resistance of the diode when this valve is conducting, the fraction of input voltage developed across the diode during the positive half-cycles is small. If R_s is small compared with the impedance of the diode when this valve is not conducting, a large fraction of the input voltage is developed across the diode during the negative half-cycles. The diode may have a resistance of between 1000 and 2000 ohms when conducting, this resistance being the ratio of the anode-cathode voltage to the current through the valve, and not the slop resistance. The impedance of a diode, if it is not conducting, is entirely due to the anode-cathode interelectrode capacitance, and this capacitance may be about 10 pF. Fig. 473 shows the limiting of the positive half-cycles in the voltage developed across the diode, and also the limiting of the negative half-cycles in the voltage developed across the resistor R_s .

Fig. 473a shows the diode limiting circuit arranged so as to obtain the output in the form of negative pulses (Fig. 473c), while Fig. 474 shows a circuit suitable for producing positive pulses (Fig. 473d). If the diode is reversed as in Fig. 475, the positive half-cycles appear mainly across the diode whilst the negative half-cycles are developed mainly across the resistor.

So far the limiting of either the positive or the negative portions of the input has been considered as beginning when the input voltage passes through zero. If, however, the cathode of the diode is held at a fixed positive voltage, as shown in Fig. 476a, the diode does not conduct until the input reaches this level, and limiting of the voltage across the diode thus occurs at a fixed positive voltage. The voltage across R_s in Fig. 476a consists of the peaks of the positive half-cycles of the input, shown dotted in Fig. 476b. Fig. 477 shows the limiting circuit in a form

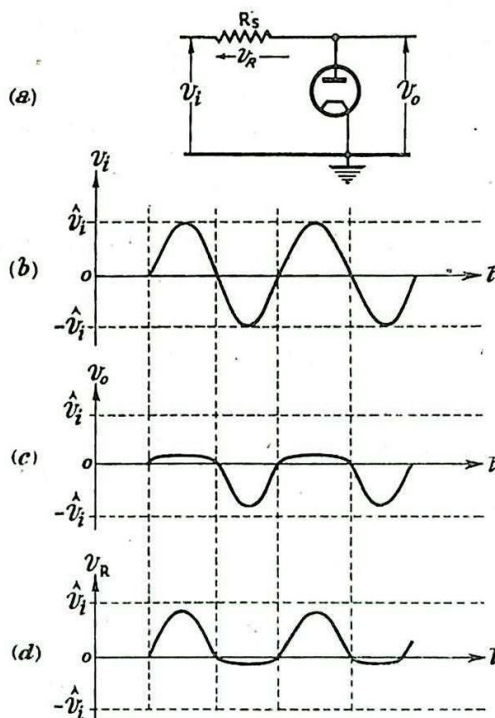


Fig. 473 - Positive peak limiting; output across diode.

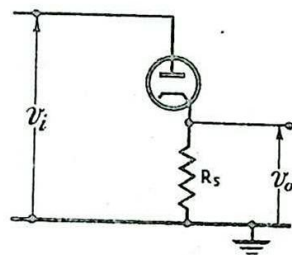


Fig. 474 - Negative peak limiting; output across resistor.

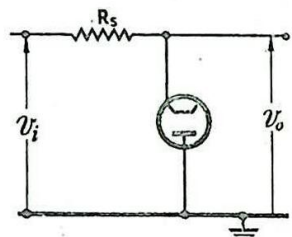


Fig. 475 - Negative peak limiting; output across diode.

suitable for obtaining this voltage as an output. In this circuit R_1 and R_2 , which are effectively in parallel, take the place of R_s in the previous circuits. Fig. 478 shows circuit arrangements in which the negative peaks are limited and the output voltage is developed across the diode. In Fig. 478b C is a blocking condenser necessitated by the change in voltage level. In both circuits the voltage across R_s consists of the peaks of the negative half-cycles of the input voltage. Figs. 479a and b show the corresponding circuits in a form suitable for obtaining this voltage across R_s as an output.

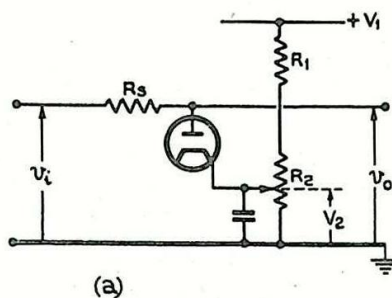
In any limiting circuit using a coupling condenser, as for example in Figs. 478b, 479b the time-constant of the input circuit must normally be chosen so that no appreciable slide-back occurs. (See Chap. 7, Sec. 4).

If a suitable choice of condenser and resistor values is made slide-back biasing circuits may be used to provide bias in some cases.

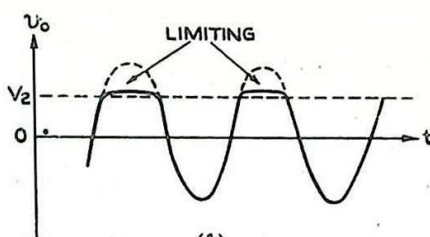
By using a pair of diodes arranged as in Fig. 480, both upper and lower excursions of the input voltage can be limited. In this circuit diode 2 is conducting if there is no input voltage, so that the common cathode voltage is

$$V_K = \frac{V R_2}{R_1 + R_2} \quad (\text{neglecting the resistance of the diode compared with } R_1 \text{ and } R_2)$$

When the input voltage reaches this value diode 1 conducts, and the



(a)



(b)

Fig. 476 - Positive peak limiting with bias; output across diode.

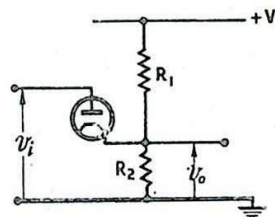


Fig. 477 - Negative peak limiting; output across resistor.

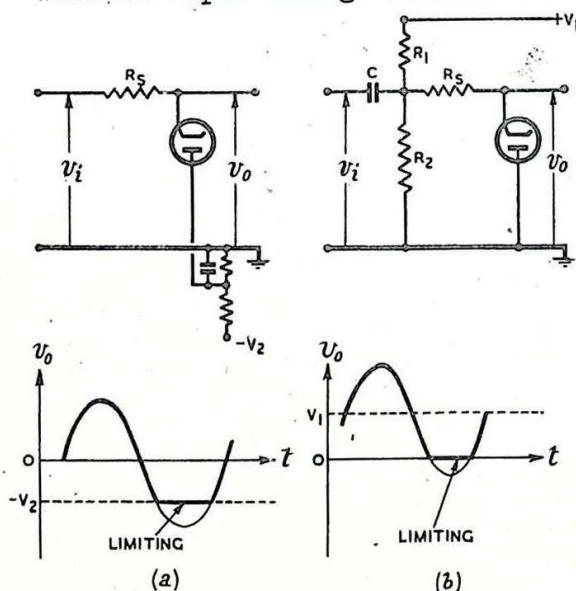


Fig. 478 - Negative peak limiting; output across diode. With no voltage shift (a) and with voltage shift (b).

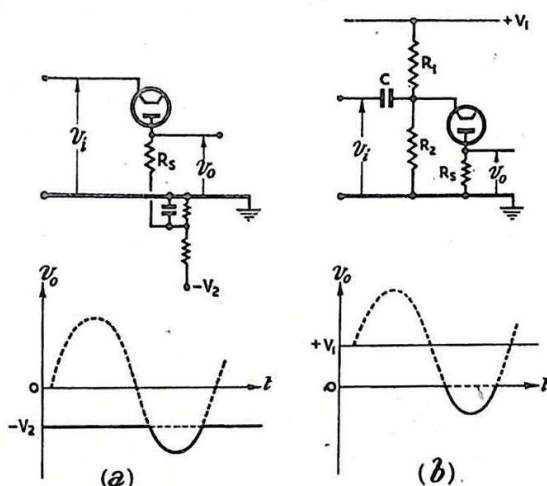


Fig. 479 - Positive peak limiting; output across resistor. With no voltage shift (a) and with voltage shift (b).

voltage developed across R_2 rises. As long as diode 2 is conducting the voltage at its anode is practically equal to the common cathode voltage. Therefore, as the voltage across R_2 rises from

$$\frac{V R_2}{R_1 + R_2}$$

the output voltage rises with it.

As soon as the voltage across R_2 rises above V , diode 2 ceases to conduct and the output voltage remains at the value V . It follows that the portion of the input lying between the values V and $\frac{V R_2}{R_1 + R_2}$ is the only portion which contributes to the output. In this and similar cases we shall call this portion of the input the Effective Input Voltage.

LIMITING CIRCUITS EMPLOYING TRIODES OR PENTODES

3. Limiting Amplifiers.

Since there is no fundamental difference between the action of a triode and that of a pentode used as an amplitude limiter, unless otherwise stated the discussion of triode circuits may be taken as applying equally well to those containing pentodes.

Consider a sinusoidal voltage applied to the control grid of a resistance-loaded amplifier. The cut-off voltage is determined by the HT supply V_B , and for values of the input voltage below cut-off the anode voltage remains constant at its maximum value V_B . By adjustment of the bias, the lower level of the input at which limiting takes place may be fixed at any desired value.

The upper level of the input at which limiting occurs depends on the position of the flat portion of the dynamic characteristic (A in Fig.). This in turn is determined by V_B and the anode load resistor. If a sufficiently large load resistance is used the anode current reaches its maximum for a correspondingly low grid voltage. This is usually advisable because of the undesirable effects of driving the grid voltage far beyond the region at which grid current begins to flow.

Since it is not usually advisable to employ a very large anode load this method of limiting is normally confined to pentodes, which do not require the large load resistances needed by triodes (see Sec. 4).

An alternative method of achieving the limiting action for the positive portions of the input is to use the grid-cathode portion of the valve as a diode. The circuit arrangement is shown in Fig. 481. The mode of operation is almost identical with that already described for the diode circuit of Fig. 473. In the circuit of Fig. 481 as long as the input voltage does not exceed the cathode voltage almost the whole value of the input voltage is developed between grid and cathode. However, as this voltage reaches the cathode voltage grid current flows and the resistance of the grid-cathode path drops to a low value, of the order of 1000 ohms. If the limiting resistor R_3 is large compared with 1000 ohms, e.g. 50 k Ω , only a very small fraction of the applied

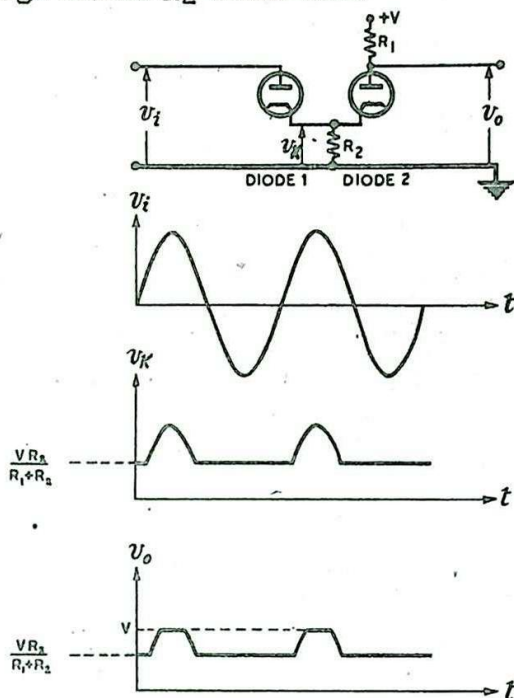


Fig. 480 - Positive and negative peak limiting using twin diodes.

voltage appears between the grid and cathode when grid current flows. The greater the value of the grid resistor the greater the limiting action. There are, however, reasons why R_g should not be too large. R_g and the input capacitance of the valve form a series circuit across the input, and it is the voltage developed across the capacitance which is actually applied to valve. The allowable maximum value of R_g is determined by the frequency f of the input voltage, since the reactance of the input capacitance of the valve must be large compared with R_g if a large fraction of the applied voltage is to appear between grid and cathode. This condition may be written $C_i R_g \ll \frac{1}{2\pi f}$, so that a short time-constant circuit is required. If the voltage appearing between grid and cathode is not a large fraction of the input voltage it may not raise or lower the grid potential sufficiently to cause either grid-current or cut-off limiting.

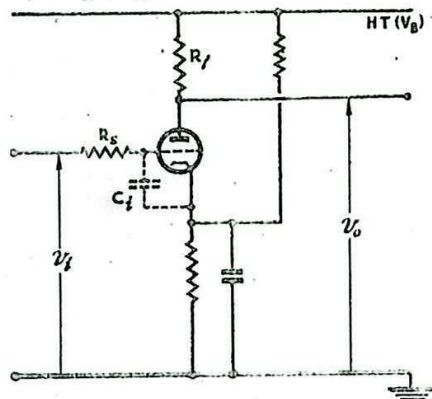


Fig. 4.81 - Circuit of triode amplitude limiter (grid current limiting).

If instead of a sinusoidal voltage a rapid change of voltage is applied to the circuit, the rapidity with which the grid voltage follows the input is determined by the time-constant $C_i R_g$. If this is too large the action of the amplifier in producing a rapid change in output voltage is largely nullified.

In the foregoing discussion the resistor R_g has been included for the specific purpose of limiting the output voltage. A similar effect is obtained if the output resistance of the generator is large. In this case the effective limiting resistance is the sum of this output resistance and any additional resistance R_g included to ensure the limiting action.

It should be noted that in the grid-current method, limiting occurs actually at the grid and is reproduced in the anode circuit. In the two other methods the grid voltage is substantially the same as the input voltage.

Almost any resistance-loaded amplifier circuit may be adapted for use as an amplitude limiter by arranging for the input voltage to operate over one or both of the flat portions of the dynamic characteristic, or by the use of a grid limiting resistor. This normally entails a suitable choice of grid base and bias. Possible arrangements for providing either positive or negative bias are described in Chap. 7, Sec. 4.

4. Relative Merits of Triodes and Pentodes.

A pentode has the following advantages over a triode for use as a limiting amplifier:-

- (i) The input capacitance of a pentode is normally smaller than that of a triode owing to the diminished Miller effect. This enables a larger limiting resistance to be used. (See Sec. 3)
- (ii) Because a higher gain can be obtained from a pentode the fall in anode voltage as the grid voltage rises through cut-off is much sharper than for a triode, so that the limiting effect at cut-off is more definite.

- (iii) The grid-base of a pentode is usually smaller than that of a triode under similar operating conditions. This makes it more suitable for the production of rectangular pulses from a sinusoidal input voltage (see Sec. 6) and for similar pulse-shaping requirements.
- (iv) Good limiting action at the flat portion of the dynamic characteristic (A in Fig. 472) is practicable with a pentode but not normally with a triode. This enables the limiting

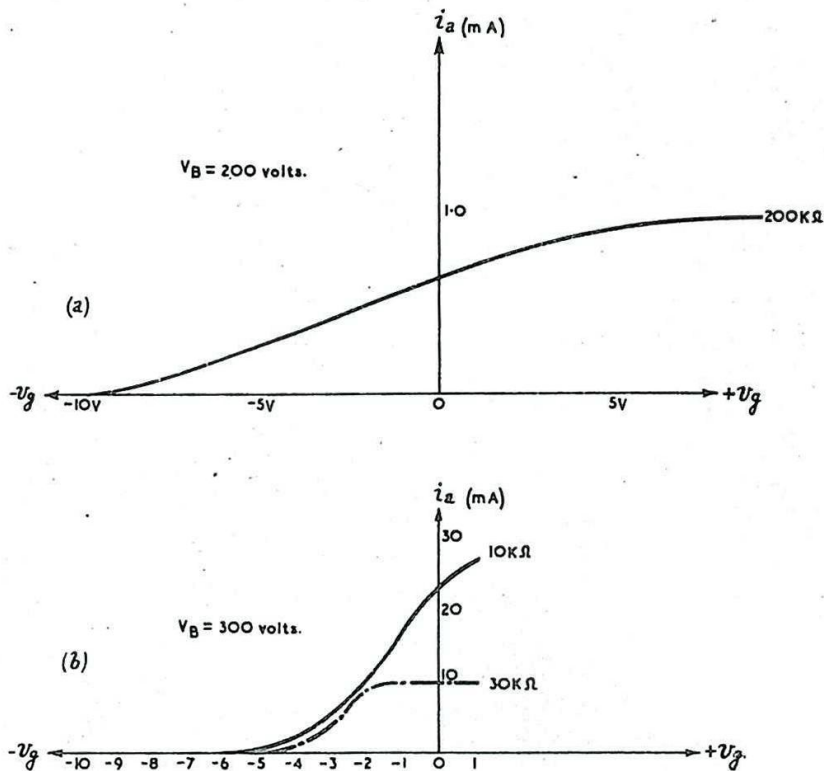


Fig. 482 - Dynamic characteristic, (a) triode, (b) pentode.

resistor to be dispensed with. Fig. 482 shows typical dynamic characteristics for a triode and a pentode. For the pentode (Fig. 482a) using a 30 kΩ anode load the limiting action occurs for grid voltages greater than that about -1.5 volts. In the case of the triode such limiting action would not occur until much higher grid voltages were reached, even with a much larger anode load and reduced HT voltage (Fig. 482b).

APPLICATIONS OF AMPLITUDE LIMITERS

5. The Production of Rectangular Pulses

Fig. 483 shows how an approximately rectangular output may be produced from a sinusoidal input by limiting action at suitably chosen levels AA' and BB'. Figs. (a)-(d) show how, by varying the bias in relation to these levels, the relative durations of the effective positive-going and negative-going portions of the input (Mark-To-Space ratio) may be controlled. If diode limiters are employed the output voltage is approximately identical with the effective input, whereas with limiting amplifiers the output voltage is an inverted and amplified version of the effective input.

The greater is the ratio of the amplitude of the sinusoidal voltage to that of the effective input the more closely does the output

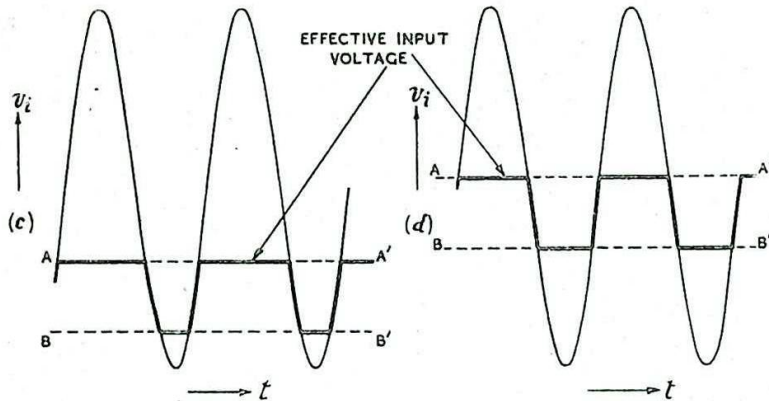
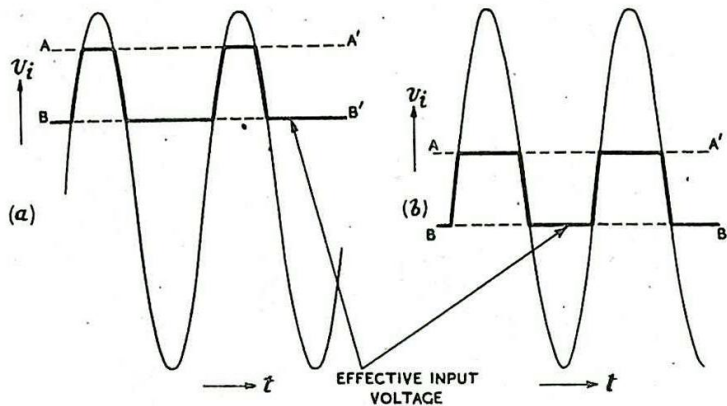


Fig. 483 - Production of "rectangular" pulses from sinusoidal input.

conform to a rectangular shape. Further, the rate of change of the sinusoidal input voltage is greatest at the points where this voltage passes through its mean value; accordingly, the more closely the effective input voltage approaches to this mean value the more nearly rectangular is the output. For example, in Fig. 483 the output voltage is more nearly rectangular in cases (b) and (d) than in cases (a) and (c).

We deduce also that as the inequality between the positive-going and negative-going portions of the output becomes more marked the output pulses deviate further from a rectangular shape.

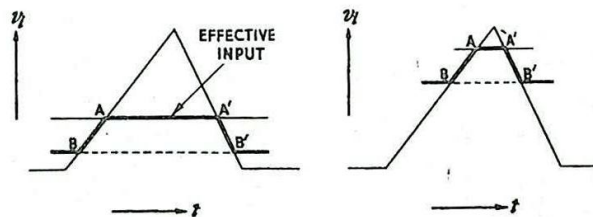


Fig. 484 - Production of "rectangular" pulses from triangular input voltage.

Fig. 484 illustrates the production of output pulses which are approximately rectangular from a triangular input. The duration of the output pulses depends upon which portions of the triangular pulses are limited.

We have assumed so far that the amplitude limiter produces an undistorted version of the effective input voltage. This is true in the case of a resistive-loaded valve, but in practice the presence of capacitance in parallel with the load may introduce considerable distortion. The effect of this on limiting amplifiers is considered in Sec. 8.

6. Elimination of Irregularities

In general the pulses encountered in practice rarely possess the perfect shape shown in Fig.485a; typical irregularities are illustrated at (b), (c), (d) and (e). If any of these pulses is applied to an amplitude limiter so that the effective input voltage lies between the limits AA' and BB', the irregularities are not present in the output.

In a similar manner, if a succession of pulses of varying amplitude (Fig.486) is applied to a limiter, the upper limiting level AA' can be fixed so that the output consists of a succession of pulses of uniform amplitude.

7. Pulse Selection

As an extreme case of the elimination of irregularities, consider the waveform of Fig.487. By a suitable choice of the upper limiting level AA' positive-going peaks of the input can be made ineffective.

In the particular example shown in Fig. 487, in which a limiting amplifier is used, not only the positive-going portions of the input voltage, but also the peaks of the negative-going portions are made ineffective.

Further examples of discrimination between wanted and unwanted pulses are shown in Figs.488 and 489.

8. TIME-CONSTANT CONSIDERATIONS AFFECTING THE OUTPUT CIRCUIT OF A LIMITING AMPLIFIER.

It has been shown in Chap. 7 Sec. 11 that capacitance in parallel with the anode load of a resistance-loaded amplifier may considerably modify the output voltage. This particularly applies to limiting amplifiers used for the production of rectangular pulses.

The circuit diagram of Fig.490 with its simple equivalent circuit (b) shows that the time-constant of the output circuit is

$$C \cdot \frac{R_a R_l}{R_a + R_l}$$

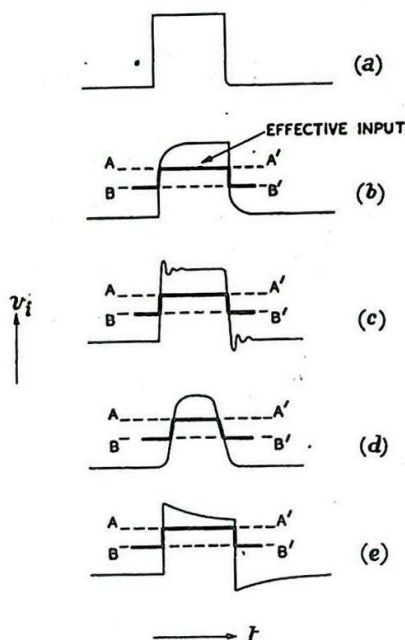


Fig.485 - Elimination of irregularities.

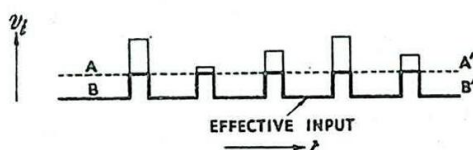


Fig.486 - Limiting the amplitude of a succession of pulses.

where R_a is the slope resistance of the valve. For a linear amplifier R_a is constant; but for a limiting amplifier which works over the curved portions of the dynamic characteristic R_a is by no means constant, and when the valve is cut off R_a is infinite. It follows that as the input rises through cut-off the time constant of the output becomes:-

$$\frac{C \cdot R_a R_f}{R_a + R_f}$$

whilst, as the input voltage falls through cut-off, the time-constant becomes

$$C R_f \text{ (Since } R_a = \infty \text{)}$$

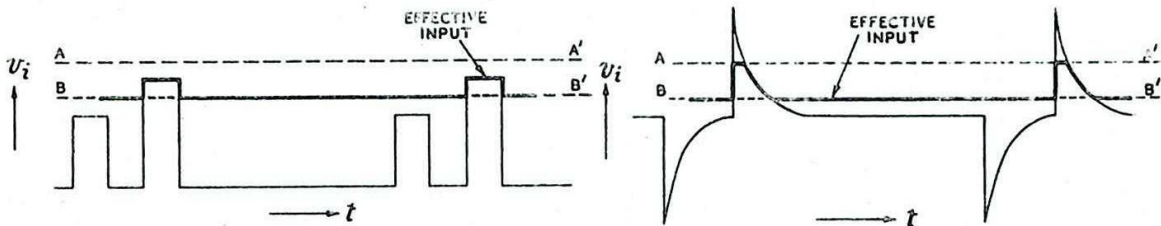


Fig.487 - Pulse selection; positive input pulses ineffective.
Fig.488 - Pulse selection; negative-ineffective below given amplitude going pulses ineffective.

Consider the response of an amplifier to a rectangular input pulse

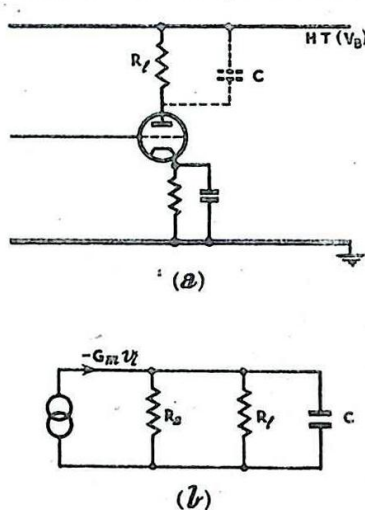


Fig.490 - Limiting amplifier circuit; time-constant considerations.

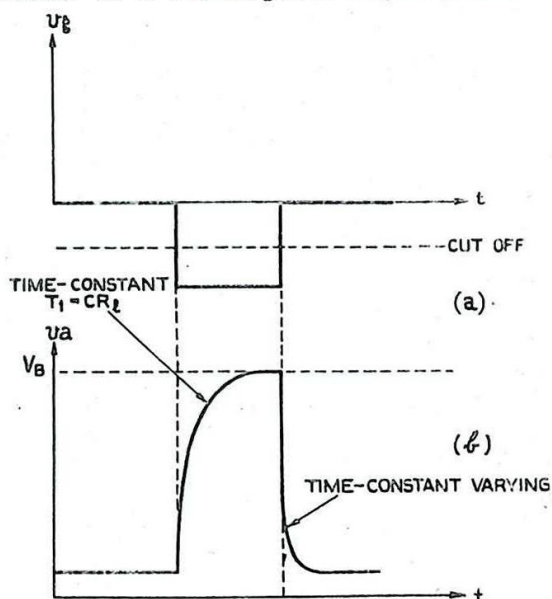


Fig.491 - Limiting amplifier; time-constant considerations.

which cuts off the valve (Fig. 491.). The initial rise of anode voltage follows an exponential law with time-constant $T_1 = CR_0$.

After the input voltage has risen through cut-off at the end of the pulse and acquired its steady value, the time constant for the fall in anode voltage is.

$$T = \frac{C R_a R_0}{R_a + R_0}$$

Since R_a is not constant this time-constant is continually changing so that the fall in voltage is not exponential. However, the time-constant is always smaller than CR_0 so that the fall in anode voltage takes place in a shorter time than the initial rise (Fig. 491.).

RELAXATION OSCILLATORS AND RELAYS1. INTRODUCTION

The circuits which are considered in this chapter involve the determination of time-intervals. In some cases (Relaxation Oscillators) their function is much the same as that of a simple oscillator, i.e. to produce recurrent changes of voltage at regular intervals of time. In others (Relays) it is required to produce, in response to some input pulse, an output variation which occurs at some predetermined instant after the occurrence of the input. In such cases the circuit acts in the nature of a delay device, and does not produce any output until an input or triggering pulse is applied.

Relaxation oscillators are normally used to generate low frequency oscillations, e.g. from a few cycles per second to a few kilocycles per second. They are widely used to control the timing of the operation of various circuits in radar equipments such as time-base generators, modulators, aerial switching valves, etc. A relaxation oscillator may readily be synchronised with a succession of pulses or any recurrent voltage variation whose frequency is slightly greater than the natural frequency of the oscillator.

Relays are used to produce delayed pulses, and find common application in time-base circuits, e.g. for the production of range markers, trace-brightening pulses, etc., at predetermined positions on the time base.

Any of these circuits may also be arranged so that when triggering or synchronising pulses are applied at regular intervals, they are in effect divided into groups of two, three or more pulses. Only the first pulse of each group affects the circuit, so that the frequency of the output voltage variations is a submultiple of the frequency of the input pulses. This process is called Counting-Down.

Each of these circuits depends for its action on the existence of at least two distinct conditions of operation, either of which may be stable or unstable. For example, in a two valve circuit either valve may be conducting while the other is non-conducting. The transition from one condition to the other takes the form of a rapid cumulative action due to regeneration inherent in the circuit. The circuit is in a stable state if an external impulse is needed to cause this transition; otherwise it remains in its existing state. The circuit is in an unstable state if, after an interval (time of relaxation) it reverts of its own accord to the alternative condition.

Such circuits may be classified under three headings, as follows:-

(i) Circuits with no stable state of operation

Circuits of this type are known as Relaxation Oscillators. A circuit of a Relaxation Oscillator is shown in Fig. 492. The condenser C charges through the resistor R until the voltage across it is sufficient to cause the neon valve to strike. (The interval during which the condenser charges is the time of relaxation). Upon striking, the valve presents a low resistance and the condenser discharges rapidly until the neon current falls below its extinction value. The valve subsequently presents an open circuit and the condenser

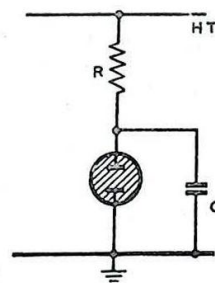


Fig.492 - Simple relaxation oscillator (flashing neon).

recharges and the cycle of operation repeats. Other well-known examples are the Multivibrator and Blocking Oscillator; these are considered below.

(ii) Circuits with one stable state of operation

A circuit of this type is known as a Relay. Such a circuit rests naturally in its stable condition, but on the application of a triggering impulse an abrupt transition to the unstable state occurs. After an interval of relaxation the circuit reverts to its stable condition.

A circuit of a relay is shown in Fig. 493. In the stable state the gas-filled triode is non-conducting, the grid voltage being maintained below cut-off and the anode voltage remaining at H.T. If the grid voltage is raised momentarily above cut-off, by an external impulse of very short duration, the valve conducts heavily and the condenser is discharged. This unstable state continues after the triggering pulse is removed, and is terminated only when the valve current falls below its extinction value. Subsequently C charges through R and the circuit reverts to its stable condition.

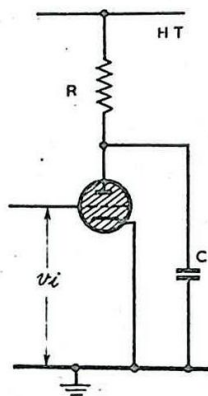


Fig. 493 - Simple relay (triggered thyratron time-base).

(iii) Circuits with two stable states of operation

A circuit of this type is known as a Counter. Such a circuit rests naturally in either of its two stable conditions, but on the application of a triggering impulse an abrupt transition from one stable state to the other takes place. The circuit rests in this second condition until a further triggering impulse causes it to return to the first. This type of circuit, a typical example of which is the Eccles-Jordan Circuit, has little if any application in radar equipment and is not discussed further.

Relaxation oscillators and relays may take the form of two-valve or single-valve circuits. In most cases the time intervals are determined by the exponential charging or discharging of a condenser through a resistor. In some circuits the linear charging principle of the Miller time base is used (see Chap. 11, Sec. 10). Circuits which utilise this principle are discussed separately in Secs. 9 to 11.

In the description of the Multivibrator circuit, which is considered first, the action is discussed more or less in detail, each main or subsidiary feature of the oscillations being dealt with as it occurs during the sequence of operation. This procedure is adopted for the sake of readers encountering this type of action for the first time. In the subsequent description of circuits where the action is similar to that of the multivibrator the main sequence of operation is described first, where convenient in note form, and any subsidiary characteristics important for an understanding of the action, or because of particular applications of the circuits, are dealt with afterwards.

It is felt that this method of description is more readily followed once the essential principles of cumulative action and

relaxation have been grasped.

TWO VALVE CIRCUITS

2. The Multivibrator

The circuit of a Multivibrator is shown in Fig. 494. This relaxation oscillator is a two-stage resistance-loaded, capacitance-coupled amplifier with the output coupled back to the input.

The pulses produced at the anodes of the two valves are approximately rectangular (Fig. 495). In other words the oscillations produced are rich in harmonics, a fact which accounts for the name "Multivibrator".

When the circuit is first switched on a transient state follows in which both valves are conducting. After this transient period the circuit will settle down to continuous oscillations as illustrated by the waveforms of Fig. 495. Consider the interval marked (1), in which the grid voltage of valve 1 is returning towards earth. As the voltage of $1G$ rises through cut-off, the flow of current in valve 1 causes a drop of voltage at $1A$ which is applied to $2G$ through C_1 . The resulting rise of voltage at $2A$ is applied back to $1G$ and there is a further increase of current in valve 1. This process is cumulative so that the voltage of $1G$ rises rapidly to a positive value and the voltage of $2G$ falls rapidly to a value which is more negative than that required to cut off the current in valve 2. Then the amplification (cumulative action) ceases.

As the voltage of $1G$ rises above zero, grid current flows, so that the grid-cathode resistance of valve 1 becomes small.

This causes the voltage rise at $1G$ to be much smaller than the corresponding fall at $2G$. (See Chap. 9 Sec. 3). The voltage at $1A$ falls momentarily below the steady value which it ultimately attains during the interval (2) because of the relatively large voltage developed initially at $1G$.

After amplification has ceased the circuit relaxes, C_2 charging and C_1 discharging. The flow of grid current in valve 1 into C_2 causes the voltage at $1G$ to fall exponentially towards zero, with a time-constant approximately equal to $R_2 C_2$, while the voltage at $2A$ rises towards V_B with the same time-constant.

Meanwhile the voltage at $2G$ rises exponentially towards zero, as C_1 discharges with a time-constant which is effectively equal to $C_1 R_4$ (providing $R_4 \gg R_1$).

As the voltage at $2G$ rises through cut-off a cumulative action ensues similar to that which terminated the interval (1). This action

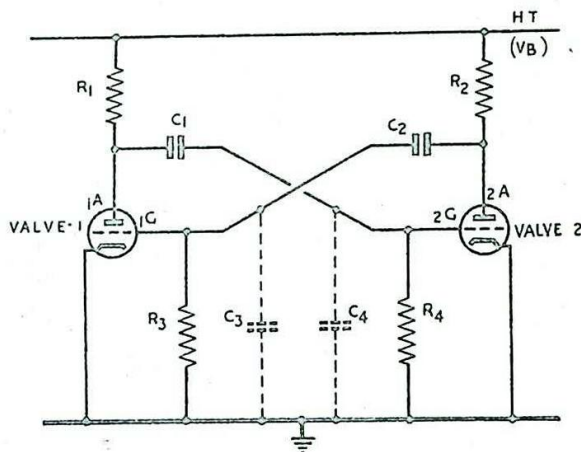


Fig. 494 - Multivibrator circuit.

is concluded when the current of valve 1 is cut off.

After amplification has ceased the circuit relaxes again (interval (3), Fig. 495). The flow of grid current in valve 2 causes the voltage at 2G to fall towards zero with a time constant which is approximately equal to $R_1 C_1$, while the voltage at 1A rises towards V_B with the same time-constant. As the voltage at 2G falls the voltage at 2A rises. The voltage at 1G rises exponentially towards zero with a time-constant which is approximately equal to $C_2 R_1$, and as this voltage rises through cut-off, valve 1 reconducts. The cycle of operation is now completed and the whole sequence repeats.

So far it has been assumed that the sudden changes in potential take place instantaneously. However, the presence of the inter-electrode and stray capacitances which are represented in Fig. 494 by C_3 and C_4 prevent any instantaneous changes. Thus, the

cumulative rise of potential at the anode of valve 1 takes place with a time-constant $R_1 C_4$, whilst that at the anode of valve 2 takes place with a time-constant $R_2 C_3$. After the cumulative action is over the anode voltage of valve 1 rises during interval (3) with a time-constant $C_1 R_1$ whilst that of valve 2 rises during interval (2) with a time-constant $C_2 R_2$. In order that these rises of potential should be rapid the anode load resistances should not be too large. The rise of anode voltage may then take place in less than 1 microsecond if use is made of receiving valves of large mutual conductance. If extremely rapid action is required (0.2 microsecond) beam power valves should be used, capable of passing large currents.

The durations of the various portions of the waveforms shown in Fig. 495 depend not only upon the time-constants but also upon the minimum values of the voltages at the control grids and the cut-off voltages of the two valves. If the minimum grid voltages for valves 1 and 2 are $-1\check{V}$ and $-2\check{V}$ while the cut-off voltages are $-1V_c$ and $-2V_c$ respectively it may be shown that the periodic time of the oscillations is given by:-

$$T = C_1 R_1 \log_e \frac{1\check{V}}{1V_c} + C_2 R_2 \log_e \frac{2\check{V}}{2V_c}.$$

For a typical multivibrator, in which

$$1\check{V} = 2\check{V} = 140, \text{ and}$$

$$1V_c = 2V_c = 10,$$

this reduces to:

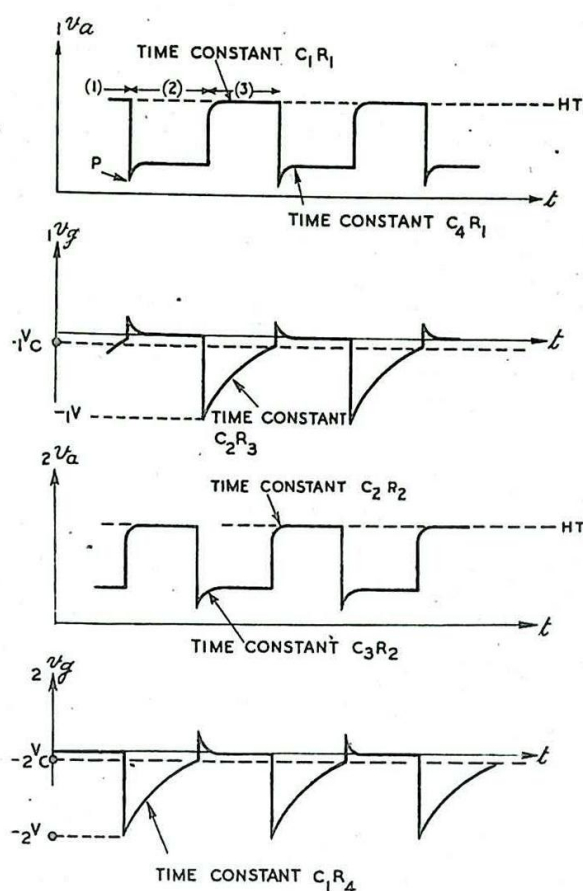


Fig. 495 - Waveforms for multivibrator circuit.

$$T \approx 2.6 (C_1 R_4 + C_2 R_3).$$

Hence the frequency is given by:

$$f \approx \frac{1}{2.6 (C_1 R_4 + C_2 R_3)}.$$

This formula is applicable only to multivibrators which operate at frequencies not greater than about 1000 c/s, since in its derivation inter-electrode and stray capacitances have been neglected.

In a symmetrical circuit the durations of the positive-going and negative-going portions of the voltage variation at either anode are equal, (Fig. 495). These portions will be of unequal duration if:-

- (i) the time-constants, $C_1 R_4$ and $C_2 R_3$ are unequal;
- (ii) valves 1 and 2 are dissimilar; or
- (iii) bias is introduced into one valve circuit only.

In radar apparatus it is sometimes necessary to change the relative durations of the positive-going and negative-going portions of the output voltage without appreciably altering the oscillation frequency. This may be accomplished, within limits, by the arrangement shown in Fig. 496. In this circuit the coupling capacitances have the same value, C . By means of a switch, the individual values R_3 and R_4 of the resistances of the two grid leaks can be altered without a change in their sum, so that the frequency:-

$$f \approx \frac{1}{2.6 C (R_3 + R_4)}$$

remains constant.

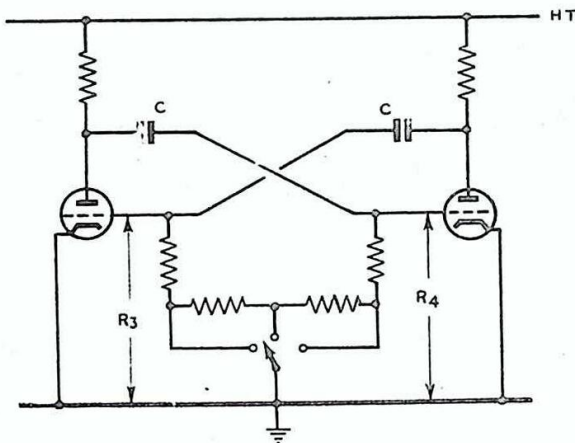


Fig. 497 - Multivibrator with grid leaks connected to HT line.

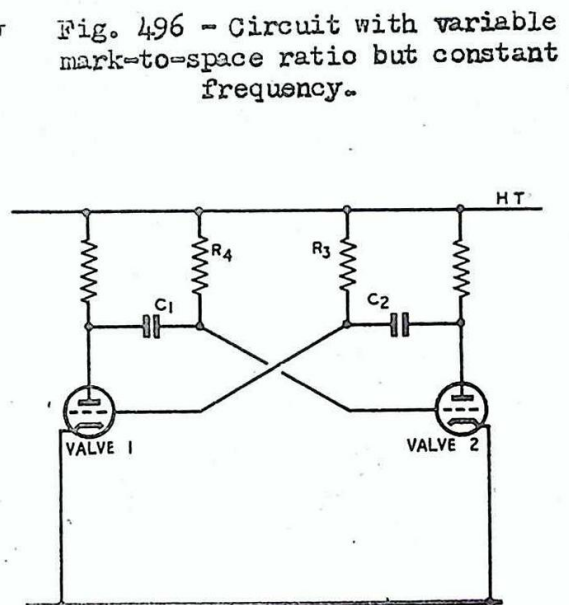


Fig. 496 - Circuit with variable mark-to-space ratio but constant frequency.

It is frequently advantageous to connect the grid leak resistors to a source of high positive potential rather than to earth; (Fig. 497). This causes the grid voltage of each valve to rise more sharply through cut-off so that the onset of the negative-going pulses at the anode is more clearly defined. In the circuit of Fig. 497, the time-constants $C_1 R_4$ and $C_2 R_3$ must be greater than those for the circuit of Fig. 494 if the same frequency of oscillation is desired.

If a succession of pulses, or any recurrent voltage variation, is introduced from an external source into the circuit of a multivibrator, the frequency of oscillation tends to increase until it is equal to, or is a multiple or submultiple of the frequency of the injected signal. Suppose that positive-going pulses are applied to the grid of valve 1 (Fig. 494), their frequency being slightly

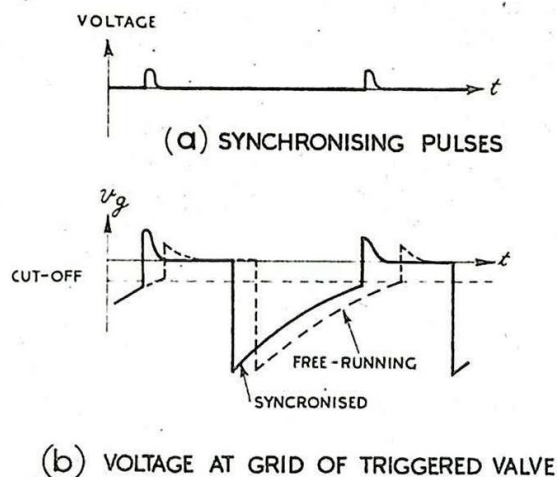


Fig. 498 - Synchronisation of a multivibrator.

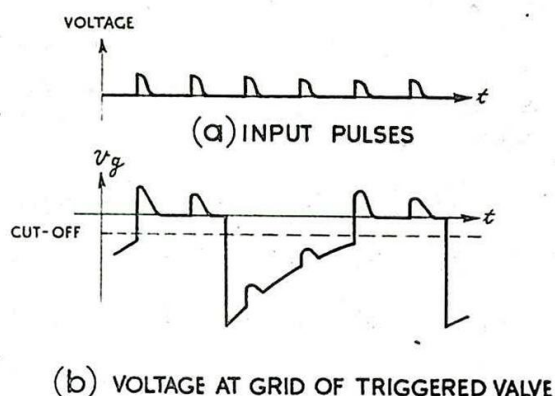


Fig. 499 - Counting-down.

greater than that of the freely-running multivibrator. Fig. 498 shows the resulting variation of potential at grid $1G$. When the pulse arrives $1G$ has not quite reached its cut-off voltage, but the application of the pulse carries it over this point, so hastening the transition to the conducting state. The multivibrator action is thus synchronised with the applied pulses. Fig. 499 shows the use of a multivibrator as a frequency divider. By a suitable choice of component values the circuit may be arranged so that it is triggered by each n -th pulse; the multivibrator frequency is then $\frac{1}{n}$ -th of the input frequency. In practice n is not usually greater than 10, although with care in adjustment it may be made as high as 50.

If only one valve is triggered in this way the duration of only the positive-going portion of its anode voltage is affected. If it is desired to control the durations of both portions of the output voltage, both valves should be triggered, e.g. by applying negative triggering pulses between the common cathode and earth.

There are several advantages in using pentodes instead of triodes in a multivibrator circuit. The presence of two additional electrodes in each valve to which pulses can be applied, or from which they can be taken, makes for greater flexibility in circuit design, and an immediate advantage is that one grid can be used exclusively for synchronising purposes. The use of pentodes also makes it possible to eliminate the peak occurring at the beginning of each negative-going portion of the anode voltage of a multivibrator (P, Fig. 495). The effectiveness of this limitation, the principle of which is discussed in Chap. 9, Sec. 4, is dependent on the anode load resistances used.

The circuit shown in Fig. 500 is that of a multivibrator using pentodes. The circuit is similar to that of Fig. 494 but the screens of the valves are used as anodes for the generation of the relaxation oscillations. Some of the space current in the conducting valve reaches the true anode and, flowing through the load resistor, provides the output voltage. In this type of circuit it is said that the anode circuits are "electron-coupled" to the remainder of the valve, as in the "Electron-Coupled" oscillator. For each valve the screen and control grid are shielded from the anode by the suppressor grid. Hence the load does not greatly affect the action of the relaxation

circuit, and the frequency of oscillation is substantially independent of changes in loading.

3. Cathode-Coupled Multivibrators

Various forms of multivibrator are in use other than the conventional type already described. Fig. 501 shows an arrangement known as a Cathode-Coupled Multivibrator. The operation of this circuit is similar to that of the conventional multivibrator and involves the same kind of

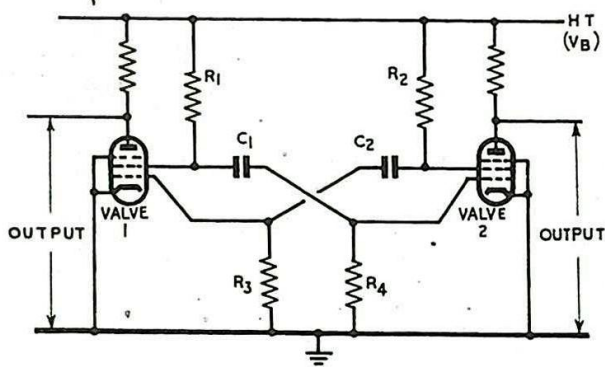


Fig. 500 - Electron-coupled multivibrator.

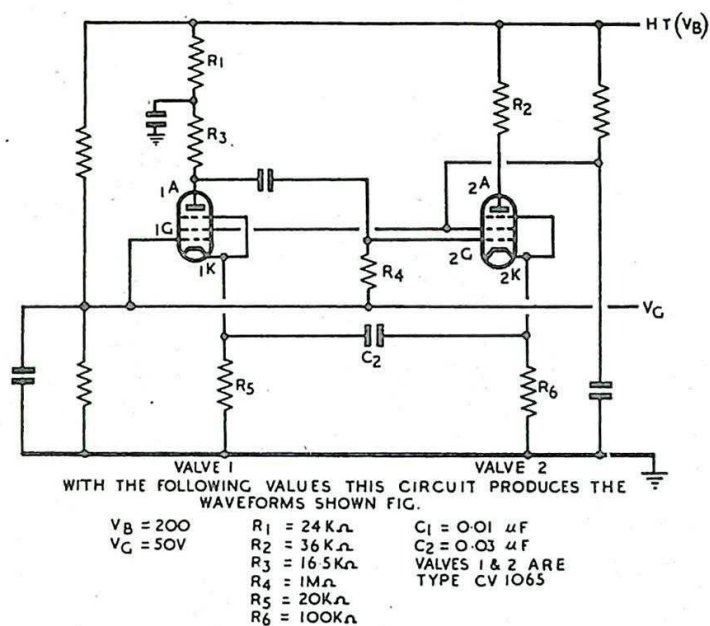


Fig. 501 - Cathode-coupled multivibrator.

502). The voltage at $1K$ falls as C_2 discharges. As the voltage at $1K$ approaches that of $1G$ valve 1 starts to conduct.

The subsequent behaviour is as follows:-

The voltage at $1A$ falls (after amplification in valve 1).
 " " " $2G$ " "
 " " " $2K$ " (as in a cathode follower).
 " " " $1K$ " still further, so that the action is cumulative.

This action is terminated when the current in valve 2 is cut off.

Interval (2)

C_2 discharges as $2K$ returns towards earth potential in the absence of current in valve 2.

As the voltage at $2K$ approaches that of $2G$ valve 2 starts to conduct.

The voltage at $2K$ rises.

cumulative action. A change in grid-cathode voltage gives rise to a chain of events resulting in a similar but amplified voltage variation being added to the voltage between grid and cathode so that the circuit is unstable.

A brief explanation of the action is given below in note form.

Assume valve 2 to be conducting and the current in valve 1 to be cut off, (interval (1) Fig.

The voltage at $1K$ rises.
 " " " $1A$ "
 (after amplification in
 valve 1).
 The voltage at $2G$ rises.
 " " " $2K$ "
 still further so that
 the action is again
 cumulative.

This action is
 terminated when the
 current in valve 1 is
 cut off.

Interval (3)

C_2 charges
 as $1K$ returns towards
 earth potential and
 the action repeats.

It has been
 assumed that no appreciable
 change occurs in the
 voltage at $2G$ during the
 intervals between the
 transitions from one state
 to the other. This
 assumption is justified
 if $C_1 R_4$ is large com-
 pared with the period of
 oscillation.

An essential
 feature of the operation
 is that the cathode load
 resistors of the two valves
 are unequal, so that when
 valve 1 is conducting $1K$
 is at a lower potential
 than $2K$ when valve 2 is
 conducting. It is the
 charging and discharging
 of C_2 through this
 potential difference
 which determines the
 period of oscillation.

The time-constant which determines the duration of interval
 (2) is approximately $C_2 R_6$, (neglecting the output resistance of the
 cathode circuit of valve 1).

The time-constant which determines the duration of interval
 (3) is approximately $C_2 R_5$ (neglecting the output resistance of the
 cathode circuit of valve 2). The frequency is thus approximately
 proportional to

$$\frac{1}{C_2 (R_5 + R_6)},$$

and the mark-to-space ratio depends on the ratio R_5/R_6 . However, if
 the values of R_5 and R_6 are altered, the amplitudes of the voltage
 variations at the cathodes are changed so that the voltage developed
 across C_2 is affected. For this reason if it is desired to change

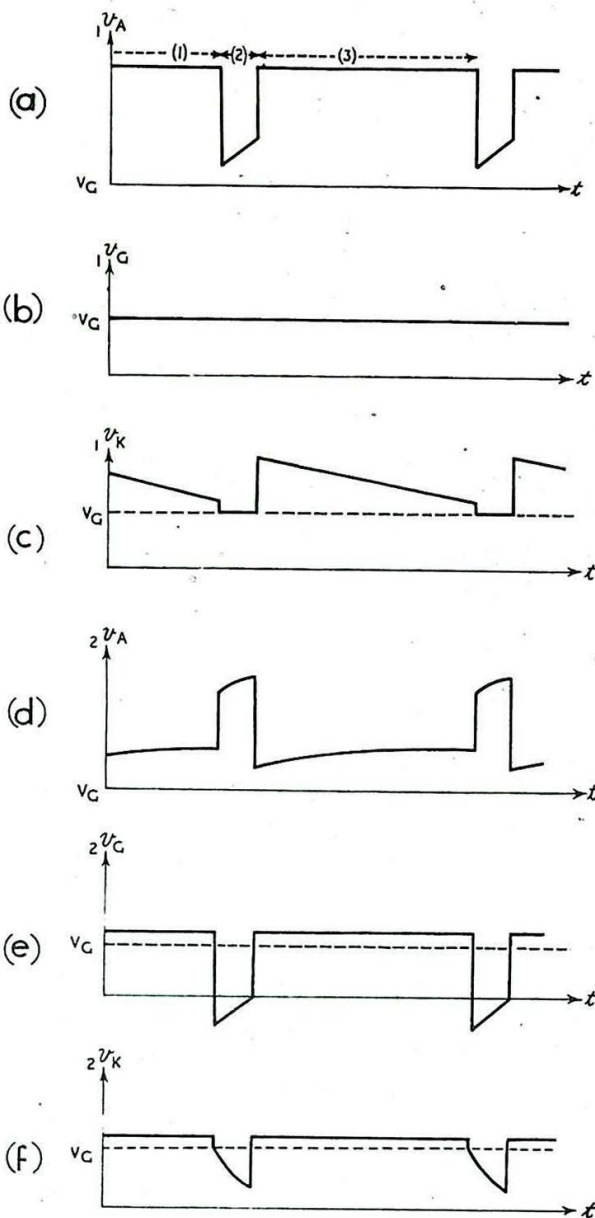


Fig 502 - Waveforms for cathode-coupled multivibrator.

the mark-to-space ratio by changing the ratio R_5/R_6 , some further adjustment must be made if the frequency of oscillation is to remain the same. This may take the form of an alteration to the value of R_3 .

Since the control grids are maintained at a high positive voltage, only a small portion is utilised of the exponential fall of either cathode towards earth. This ensures that the instants at which cumulative action occurs are clearly defined.

An alternative form of cathode-coupled multivibrator is shown in Fig. 503; the waveforms for this circuit are given in Fig. 504. The relaxation times are determined by the charging of C_2 through R_1 and the discharging of C_3 through R_3 . In this case the asymmetry in the cathode circuits is provided by the resistor R_5 .

In some cathode-coupled multivibrators of this type the asymmetry which is essential for the operation is afforded by the use of valves which are of different current-carrying capacities.

4. Relay Circuits

If the bias voltage applied to one of the valves of a multivibrator is made sufficiently negative the circuit becomes so unbalanced that it is no longer

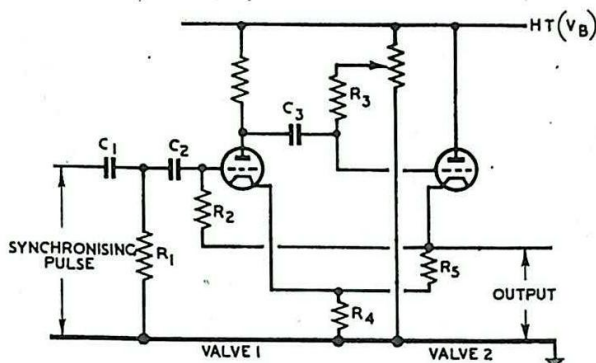


Fig. 503 - Alternative form of cathode-coupled multivibrator.

I_1 = CURRENT THROUGH VALVE 1 AT ZERO BIAS
 I_2 = CURRENT THROUGH VALVE 2 AT ZERO BIAS

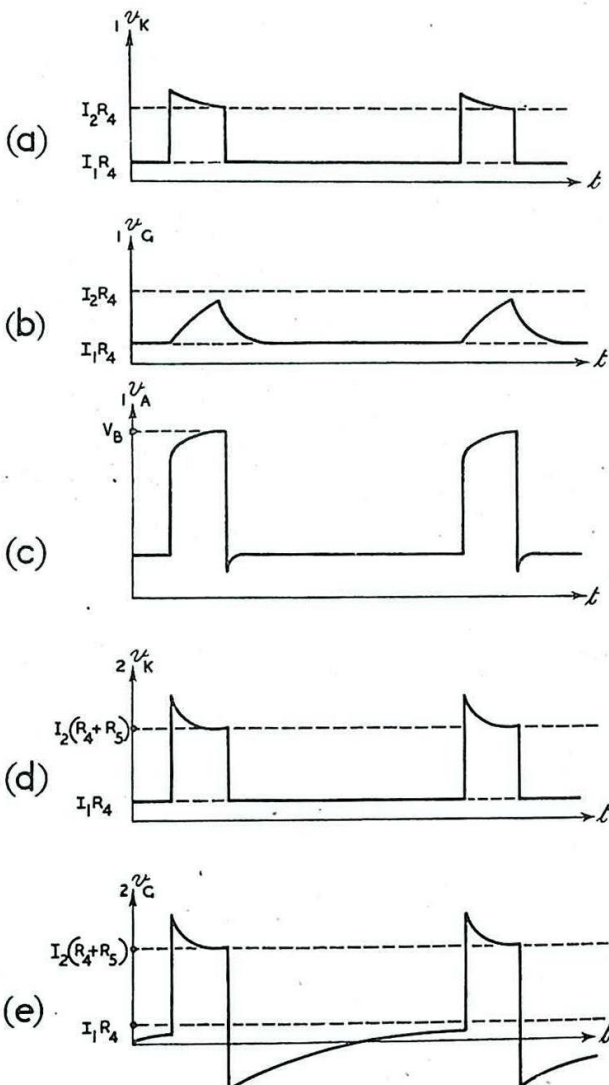


Fig. 504 - Waveforms for alternative form of cathode-coupled multivibrator.

free-running. The circuit then operates as a relay, the stable state being that in which the unbiased valve is conducting. The circuit can be triggered from the stable to the unstable state by the leading edge of an applied pulse, which causes the biased valve to conduct. The subsequent action is similar to that of a multivibrator except that after the biased valve has again become non-conducting it remains in this condition until a further triggering pulse is applied.

In general a relay is used for the production of rectangular pulses of comparatively long duration and large amplitude compared with the duration and amplitude of the applied pulses. The nature of each output pulse is normally not dependent upon the amplitude or duration of the triggering pulse, provided the latter is sufficiently large to initiate the action.

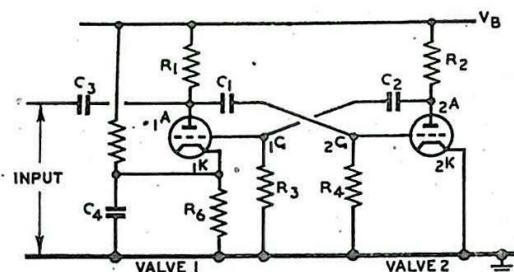


Fig. 505 - Two-valve relay.

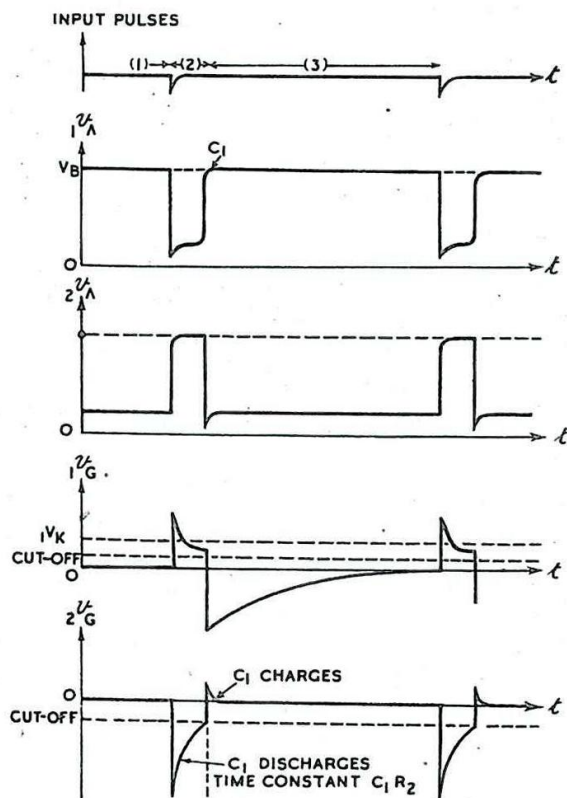


Fig. 506 - Action of two-valve relay.

Fig. 505 shows a possible circuit arrangement of a relay. The circuit is essentially that of the multivibrator, but valve 1 is biased so that the current in this valve is normally cut off, whereas valve 2 is unbiased. Initially, therefore, the anode voltage of valve 1 is at HT (V_B), while the anode voltage of valve 2 is at a low level (Interval (1) Fig. 506).

Assume that the relay is triggered by a negative-going pulse applied to the anode 1A of valve 1 (i.e., to the grid 2G of valve 2). The subsequent behaviour is as follows:-

The voltage at 2G falls.
 " " " 2A rises.
 " " " 1G rises
 by an amount sufficient
 to cause valve 1 to con-
 duct.
 The voltage at 1A falls
 still further, and the

action is thus cumulative.

This transition ceases when the current in valve 2 is cut off.

Interval (2)

C_1 discharges as 2G returns towards earth potential. As the voltage at 2G rises through cut-off, anode current flows in

valve 2:-

The voltage at 2A falls.
 " " " 1G " "
 " " " 1A rises.
 " " " 2G "

still further, and the cumulative action ceases when the current in valve 1 is cut off.

Interval (3)

C_2 discharges as $1G$ returns towards earth potential. Valve 1 remains cut off by the bias developed across R_6 .

The duration of interval (2) depends on the time-constant $C_1 R_4$ (assuming $R_4 \gg R_1$). The other time-constant $C_2 R_2$ must be sufficiently long to prevent any substantial change in the voltage at $1G$ during the interval (2), but short compared with the interval (3) so that the circuit returns to its initial state before the next triggering pulse is applied. If the latter condition is not satisfied the circuit does not respond to the next pulse, and counting-down occurs (Sec.2).

The time-constant of the discharge of C_2 can be considerably reduced by the insertion of a diode between $1G$ and earth, with its cathode connected to $1G$.

The relay may also be triggered by the application of a positive-going pulse to $2A$ or $1G$. If this method is employed a larger pulse is necessary since the amplification afforded by valve 2 is dispensed with.

The method of triggering indicated in Fig. 505 has the additional advantage that as soon as valve 1 conducts the input resistance of the anode circuit is low, so that the normal operation of the relay is not likely to be affected by the trailing edge of the triggering pulse.

Secondary effects occur similar to those which arise in the case of the multivibrator; these are indicated in the waveforms of Fig. 506.

Fig. 507 shows an alternative method of biasing. The current flowing in valve 2 develops across the common cathode resistor a bias voltage sufficiently large to ensure that the current in valve 1 is cut off in the absence of a triggering pulse.

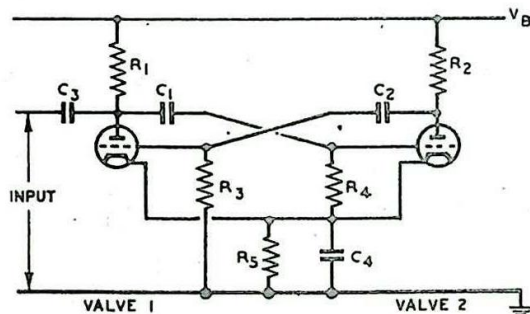


Fig. 507 - Two-valve relay with alternative biasing arrangement:

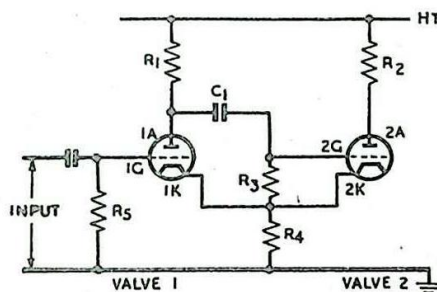


Fig. 508 - Cathode-coupled relay.

In both circuits the onset of the trailing edge of the output pulse is more clearly defined if the grid leak R_1 is connected to the HT supply instead of to the cathode (compare Sec. 2).

Fig. 508 shows the circuit of a cathode-coupled relay. The action is similar to that of a cathode-coupled multivibrator. It is essential for the relay action that the current flowing in valve 2 when conducting is greater than that in valve 1 when valve 1 is conducting. This may be ensured by using different valves, or by making $R_1 > R_2$.

The action is as follows:-

Interval (1)

Valve 1
current cut off, valve
2 conducting. A
positive triggering
pulse causes valve 1
to conduct.

Voltage at $1A$ falls,
" " $2G$ " ,
Current in valve 2 is
reduced.

The current flowing
through R_1 decreases,
since valve 1 does
not conduct as
heavily as valve 2.

Hence the voltage at
 $1K$ decreases.
Current through valve
1 increases.

The voltage at $1A$
falls still further,
so that the action
is cumulative,
ceasing when valve 2
current is cut off.

Interval (2)

C_1 dis-
charges; as the
voltage at $2G$ rises
through cut-off:-

Current flows in
valve 2.

Voltage at both
cathodes rises.

Current in valve 1 is reduced.

Voltage at $1A$ rises.

Voltage at $2G$ rises, the cumulative action ceasing when the current in valve 1 is cut off.

Interval (3)

As the voltage at $2G$ rises above that of $2K$ grid current flows and C_1 charges rapidly as the relay resumes its stable state.

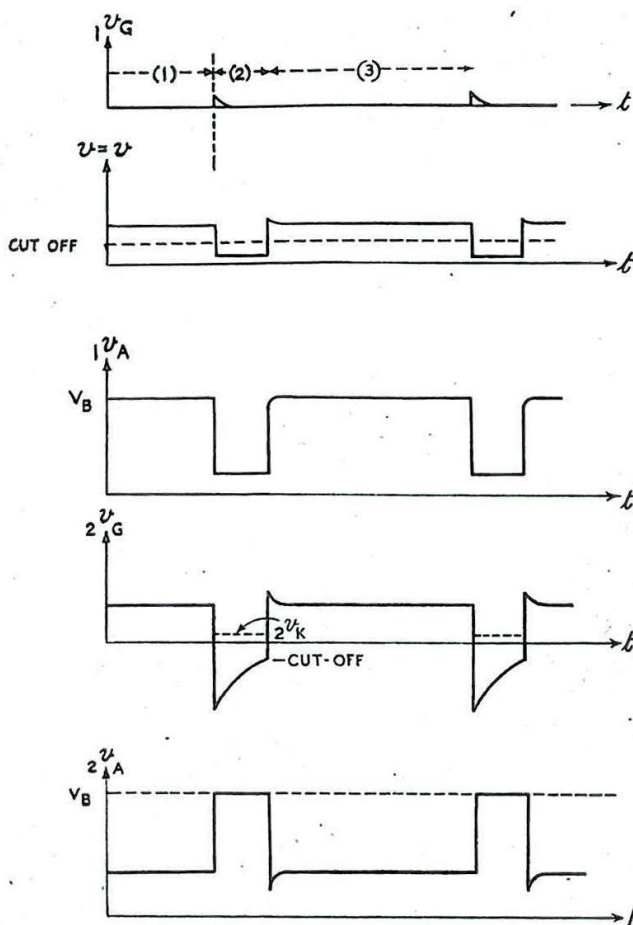


Fig. 509 - Waveforms for cathode-coupled relay.

ONE-VALVE CIRCUITS

5. Transitron Circuits

A single multi-electrode valve such as a pentode may be used to generate voltage variations by a mode of operation similar to the action of two-valve relaxation oscillators and relays. Under certain conditions the space charge between screen and suppressor grids of a pentode can act as a virtual cathode, so that variation of the suppressor grid potential can be used to control the anode current. In these circumstances, this virtual cathode, together with the anode and suppressor grid, can operate as one triode valve, whilst the true cathode, control grid and screen grid act as another. In this manner one valve may be made to fulfil the functions of the two valves in the circuits already described. The essential similarity between the one-valve and two-valve circuits will be seen in the circuit diagrams as soon as the principle is grasped of treating the pentode as a pair of triodes. There are, however, one or two characteristics which apply particularly to Transitrons, as one-valve circuits of this type are called:-

- (i) There is inherent in the circuit, but only for certain values of anode, suppressor and screen voltages, a mutual interaction between anode and screen grid. So long as a virtual cathode is present, a reduction in anode voltage diminishes the anode current and causes an increase in screen current. This leads to a fall in screen voltage due to the increased current through the screen load or dropping resistance. On the other hand, the converse is not true; a reduction in screen voltage diminishes both the anode and screen-grid currents, so that the anode voltage rises due to the decreased current through the anode load.
- (ii) In the two-valve circuits the two stable or unstable conditions occur when, in turn, each valve is conducting while the other is cut-off. One of the conditions arising in the transitron is similar, namely that in which the anode current is cut-off by suppressor grid action. No condition can arise, however, in which screen current is cut off, since this would imply that the anode current were cut off, too. Instead, the two states may be distinguished by the screen-voltage levels. When anode current flows, the screen current is small and the screen voltage high; when anode current is cut-off, screen current is large and the screen voltage is low. Relaxation oscillations or relay actions take the form of transitions from one of these states to the other and back again.
- (iii) As described in Chap. 6 Sec. 34, and illustrated in Fig. 300(a), for certain values of anode and suppressor voltages a rise in voltage at the suppressor grid leads to a decrease in screen current; so that if the screen load is not decoupled the screen voltage rises. An essential feature of transitron circuits is that the screen and suppressor grids are coupled by a condenser so that over this region of the valve characteristics (B to C, Fig. 300(a)) the action of the circuit is cumulative. A rise in suppressor voltage causes a rise in screen voltage due to the electronic action in the valve, and this in turn causes the suppressor voltage to rise still further due to the coupling circuit. A similar cumulative action follows a fall in suppressor voltage over the same region.

It should be borne in mind that whether this cumulative action

can occur or not depends on the potentials of the other electrodes in the valve; in particular, that of the anode needs to be low.

- (iv) If a feed-back resistor is included in the cathode lead this acts in a manner similar to that of the common cathode resistor of a cathode-coupled multivibrator or relay. When anode current flows the bias developed across this resistor is likely to be much greater than when the anode current is cut-off by suppressor grid action.

6. Transitron Relaxation Oscillator

Transitron Relaxation Oscillators are not in normal use in radar equipment since in general they are less stable, and also less flexible in design than the two-valve circuits. However, one arrangement, known as the Fleming-Williams circuit, is suitable as a free-running generator of sawtooth voltages.

The circuit arrangement is shown in Fig. 510. The wave-forms of the voltages at the various electrodes are shown in Fig. 511.

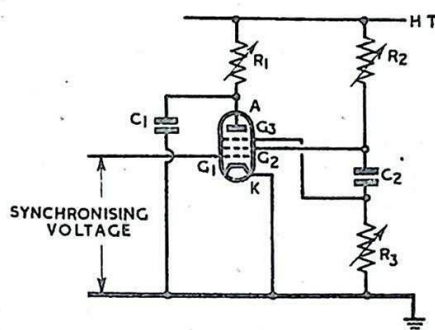
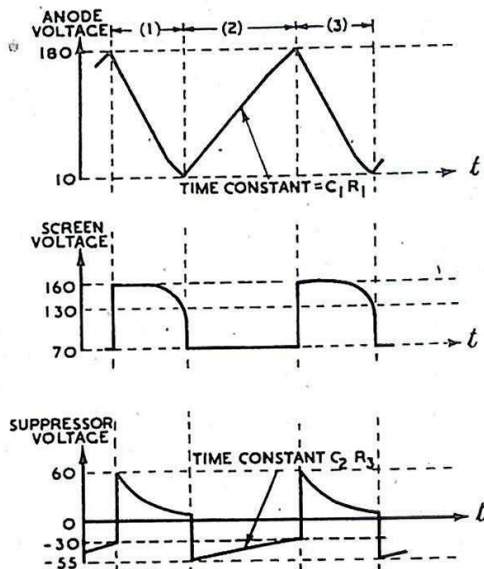


Fig. 510 = Transitron relaxation oscillator.

The action is as follows:-



VALUES OF COMPONENTS USED:-

$R_1 = 100k\Omega$ $R_2 = 500k\Omega$ $R_3 = 250k\Omega$
 $C_1 = 0.1\mu F$ $C_2 = 1\mu F$

THE VALVE IS A CV1091 WITH A HT SUPPLY OF 350 VOLTS

Fig. 511 - Action of transitron relaxation oscillator.

Interval (1)

Anode current is flowing so that the screen voltage is at the higher of its two levels of operation. C_1 discharges as the anode voltage falls. The suppressor voltage falls from a positive value. As the anode voltage falls a virtual cathode begins to form between screen and suppressor grids and a state is eventually reached at which the cumulative action described in Sec.5 (iii) is initiated. The fall in suppressor voltage causes a fall in screen voltage and because of the coupling circuit formed by C_2 and R_3 the suppressor voltage is further reduced. This cumulative action ceases when the anode current is cut off (see Fig.511).

The voltage at G_2 has then fallen to the lower of its two levels of operation (Fig. 511).

Interval (2)

C_2 discharges and the voltage at G_3 rises exponentially towards zero with a time constant approximately $C_2 R_3$ ($R_2 \ll R_3$). At the same time C_1 charges through R_1 and the anode voltage rises. A point (suppressor-grid cut-off) is reached at which anode current starts to flow, so that the screen current decreases. Hence:-

The voltage at G_2 rises.

" " " G_3 rises;

this further increases the anode current and reduces the screen current so that the action is cumulative.

This cumulative action ceases when the virtual cathode between G_2 and G_3 disappears (the anode voltage is now high), so that a further rise in the potential of G_3 has negligible effect on the anode current. This occurs when the voltage at G_3 is in the region of zero potential. However, the voltage at G_3 is raised considerably above zero by the rise in screen voltage.

Interval (3)

C_2 charges and the voltage at G_3 falls towards zero. The anode voltage falls as C_1 discharges and the cycle of operation repeats.

The oscillations may be synchronised satisfactorily by the application of either positive pulses to the suppressor grid or negative pulses to the control grid. In either case the applied pulse initiates the cumulative action from interval (2) to interval (3) (Fig. 511).

7. Transitron Relay

In the circuit depicted in Fig. 512, the components are chosen so that a stable state exists in which there is no virtual cathode between screen and suppressor grids; this is done by making the anode load sufficiently small. Alternatively, a stable state can be obtained if the suppressor grid is connected to a source of voltage sufficiently negative to prevent anode current from flowing. The condenser C_1 of Fig. 510 is omitted from the relay circuit when a rectangular output voltage is required at the anode. The relay may be triggered by either a positive-going or a negative-going pulse of small amplitude applied either to the control grid G_1 or to the suppressor grid G_3 . A larger triggering pulse is needed at G_3 than will suffice at G_1 .

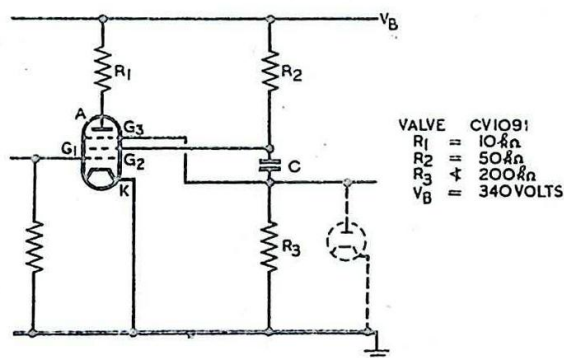


Fig. 512 - Transitron relay.

Fig. 513 illustrates the action following the application of a small positive pulse to G_1 . Initially, during the interval (1), anode current is flowing and the suppressor grid is at zero potential.

Interval (1)

As the voltage at G_1 rises:-

the space current increases, the voltage at G_2 falls, " " " G_3 falls, the anode current is reduced and the voltage at A rises.

The reduction in anode current is accompanied by a further increase in screen current causing the voltage at G_2 to fall still further.

This cumulative action ceases when the anode current is cut off so that all the space current flows to the screen.

The screen current decreases and the voltage at G_2 and G_3 rises as the input voltage at G_1 returns to zero; but provided the fall in voltage at G_2 and G_3 during the cumulative action is large enough, the anode current remains cut-off by the negative voltage at G_3 .

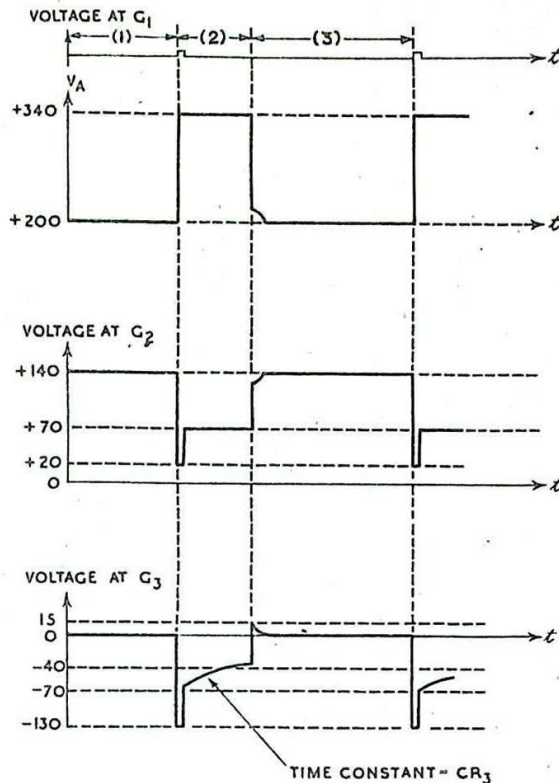


Fig. 513 - Action of transitron relay: positive triggering pulses.

Interval (2)

C discharges with time-constant approximately CR_3 , ($R_3 \gg R_2$).

The voltage at G_3 rises towards zero.

At a certain value (suppressor-grid cut-off) anode current begins to flow.

The voltage at A falls.

The increase in anode current is accompanied by a reduction in screen current causing the voltage at G_2 to rise.

This causes a further rise in the voltage at G_3 so that the action is cumulative and the relay returns to its stable condition (interval (3)).

Since the suppressor voltage rises above zero due to the rise in screen voltage, suppressor current flows and C charges rapidly.

This flow of current through R_2 causes the irregularity shown in the anode and screen-grid waveforms at the trailing edge of the output pulse (beginning of interval (3)).

The response of the circuit to a negative triggering pulse applied to screen or suppressor grid is essentially the same as in the case described above.

The effect of applying a negative triggering pulse to the control grid is illustrated in Fig. 514. No cumulative action follows the initial fall in voltage; this merely results in amplification at screen and anode as shown in the diagram, the voltage at the suppressor grid G_3 rising with the screen voltage. However, provided the triggering pulse is of sufficient amplitude or duration, the condenser C charges, due to suppressor current flowing. Due to this charging of C , the fall in screen voltage at the end of the applied pulse carries the suppressor voltage below zero potential, initiating the same cumulative action as was described with reference to the response of the relay to a positive triggering pulse. The remainder of the action is identical with that which follows in the former case.

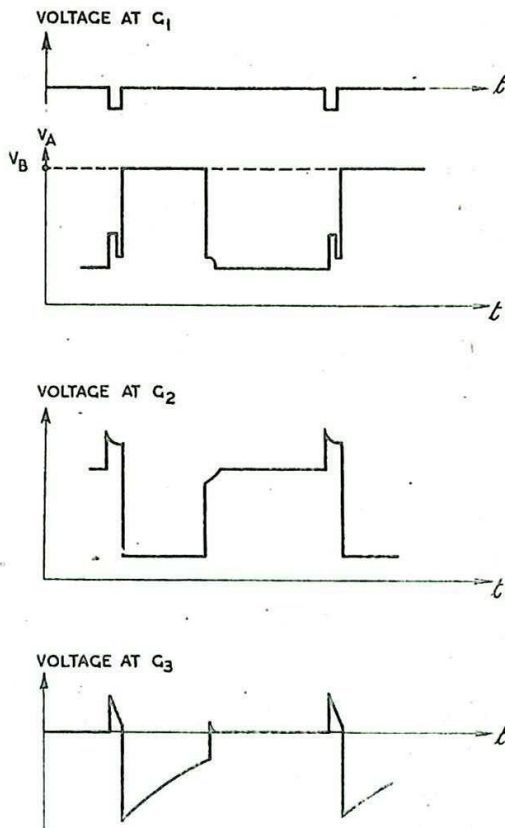


Fig. 514 - Action of transitron relay : negative triggering pulse.

If the initial rise of suppressor grid voltage is very large the effects described may be offset by the flow of reversed suppressor current.

In some pentodes (e.g. CV1091)

this may occur when the suppressor grid is about 40V above that of cathode. In such a case the net effect of the pulse applied to the control grid is to cause the condenser C to discharge rather than charge, and no cumulative action can occur. This effect can be eliminated by inserting a diode between G_3 and cathode, with the anode connected to G_3 , as shown by the dotted lines in Fig. 512. The emission of electrons from the diode cathode is likely to exceed considerably any secondary emission from the suppressor grid, so that the charging of C is ensured.

The response of the relay to a positive triggering pulse applied to the screen or suppressor grids is similar to the case just described.

In all cases the duration of the output pulse, positive-going at the anode and negative-going at the screen, depends upon the time-constant CR_3 . Owing to the large value of screen resistor necessary for the operation of the relay it is not normally practicable to take the output from the screen grid unless it is to be applied to a circuit with a correspondingly high input resistance and low input capacitance. No such limitations normally apply to the anode circuit, where the output resistance is small.

Owing to the fact that valves are not normally manufactured to give precise suppressor-grid characteristics the behaviour of a transitron circuit may vary considerably if different valves, even of the same type, are used. For this reason two-valve circuits are frequently to be preferred.

8. The Blocking Oscillator

The Blocking Oscillator is a form of relaxation oscillator

which is used primarily as a generator of rectangular pulses. The duration of the output pulses is normally of the order of 10 - 100 microseconds. The principle of the blocking oscillator is similar to that of the self-quenching oscillator, described in Chap. 8 Sec. 47. (It is common American practice to use the same term Blocking Oscillator for both types of circuit.) In the blocking oscillator the regeneration is sufficient to ensure that the valve current is out off before a single cycle is completed.

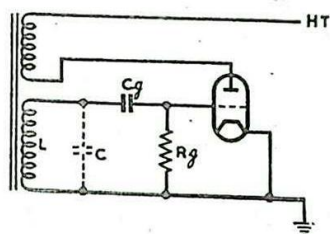


Fig. 515 - Blocking oscillator.

The circuit arrangement of one form of blocking oscillator is shown in Fig. 515. The waveforms for this circuit are given in Fig. 516.

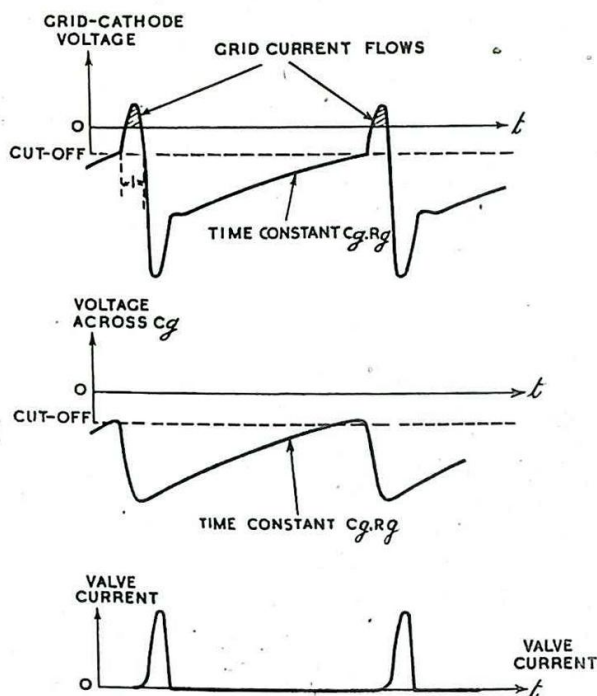


Fig. 516 - Blocking oscillator waveforms.

Although the circuit diagram is similar to that of a tuned grid oscillator the effective "tuned circuit" is much more complex. No tuning capacitance is used, and the inductance of the grid coil in parallel with its stray capacitance forms a ringing circuit. When grid current flows this circuit is heavily damped and is shunted by the series arm formed by C_g and the grid-cathode resistance. The period of oscillations in the grid circuit is thus considerably increased by the flow of grid current.

Two essential conditions of operation of the circuit are:-

- (i) The coupling between output and input circuits must be very tight, so that the grid voltage is forced to a high positive level at the beginning of the cycle (period (1)). This ensures a large flow of grid current to charge the bias condenser C_g .
- (ii) Although the time-constant $C_g R_g$ must be large, so that there is an appreciable interval of relaxation between the output pulses, C_g must be small. This ensures that the bias voltage developed across C_g acquires a large negative value before the end of the positive pulse at the grid.

If the bias is not large enough, subsequent free oscillations in the grid circuit may raise the grid voltage above cut-off again.

The relaxation time is determined by the maximum value of the bias and by the time-constant $C_g R_g$. Since the duration of the pulse is normally much shorter than the relaxation time, the latter is the main factor in determining the repetition period.

The duration of the output pulse depends on the value of the inductance and the various capacitances and resistances associated with

the grid circuit. It may be reduced by:-

- (i) reducing the effective inductance;
- (ii) reducing the capacitance C_g ;
- (iii) increasing the flow of grid current, e.g. by the use of a valve with greater cathode emission.

The grid leak may be connected to a positive supply instead of to earth. In this case the grid voltage rises towards a positive potential so that the instant at which valve current starts to flow is more precisely determined than when the grid voltage rises towards zero (see Chap. 8, Sec. 47).

The blocking oscillator may be readily synchronised by a positive-going pulse applied to the grid; alternatively, the time-constant may be adjusted so that counting-down occurs.

An alternative form of the blocking oscillator circuit is shown in Fig. 517, where an electron-coupled Hartley Oscillator is used (see Chap. 8, Sec. 8). In this case the effective inductance of the grid circuit depends, among other things, on the value of the damping resistor R , in the primary circuit. A decrease in R reduces the effective inductance and increases the damping, so that the amplitude and duration of the output pulses are reduced. The decrease in amplitude causes an increase in recurrence frequency. The recurrence frequency may also be varied more or less independently of the duration of the pulses by altering the value of the grid leak.

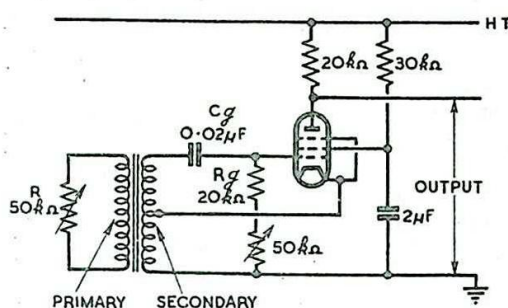


Fig. 517 - Alternative form of blocking oscillator.

With the component values shown in the figure, the duration of the negative-going pulses at the anode is of the order of 15 microseconds, and the repetition frequency can be varied from about 500 c/s. - 2000 c/s.

Blocking oscillators are frequently used with the automatic biasing network in the cathode circuit. In circuits of this type it is not practicable to use a very large resistance in the cathode lead, so that the condenser must be made larger than in the grid-biasing circuits in order to provide a long time-constant for the discharge or relaxation interval. This increase in the circuit capacitance tends to cause the output pulse to be of longer duration in the cathode-biased oscillators than in the grid-biased circuits.

More complicated circuits may be designed to make the recurrence frequency substantially independent of fluctuations in supply voltage. These arrangements generally incorporate a ringing circuit, the period of which is unaffected by changes in supply voltage. This technique constitutes a marked departure from the "relaxation" method of determining time-intervals.

CIRCUITS EMPLOYING THE MILLER TIME-BASE PRINCIPLE

9. General

The circuits discussed in the remaining sections of this chapter employ the method used in the Miller Time-Base Circuit of generating linear voltage variations. This principle is dealt with in detail in Chap. 11 Sec. 10. It involves the use of a pentode valve is controlled externally by the voltage at its suppressor grid. When this voltage is such as to allow anode current to flow, the anode voltage falls as the condenser discharges. The anode "run-down" is very nearly proportional to time. In the Miller time-base circuit some additional control device is needed for releasing and resetting the Miller valve, and also for maintaining a constant voltage at its suppressor grid during the run-down interval.

In the circuits about to be described the principle of the Miller time-base is embodied in a relay action. In the case of the Sanatron, a second valve is used for controlling the Miller valve in response to an externally applied triggering pulse. In the Phantatron circuit a single pentode valve is used to fulfil the double functions of control and Miller valves. Many variants of these circuits are used, either as relays or as relaxation oscillators, but only the two principle circuits are described here.

10. The Sanatron

A simplified circuit diagram of this relay is given in Fig. 518. An outline of the action is as follows:-

Initially valve 2 is conducting, while the anode current of valve 1 is cut off by the negative bias on its suppressor grid $1G_3$. A negative triggering pulse is applied to the suppressor grid $2G_3$ of valve 2. This is of sufficient amplitude to cut off the anode current in valve 2.

Voltage at $2A$ rises.
Voltage at $1G_3$ rises.

Thereafter, while the anode current in valve 2 remains cut off, these voltages are held at steady values by the potentiometer chain between V_B and $-V_1$.

As the voltage at $1G_3$ rises towards a positive value, anode current flows in valve 1, causing an instantaneous fall in anode voltage followed by the linear run-down. This fall in voltage at $1A$ is communicated to the control grid $2G_1$ via the network $C_2 - R_3$ and is sufficient to cut-off the current in valve 2. The rate of run-down is given, as in the Miller time-base generator, by:-

$$\frac{dV_A}{dt} = -\frac{V_2}{C_1 R_1}$$

At the end of the run-down at $1A$, the anode voltage "bottoms", C_2

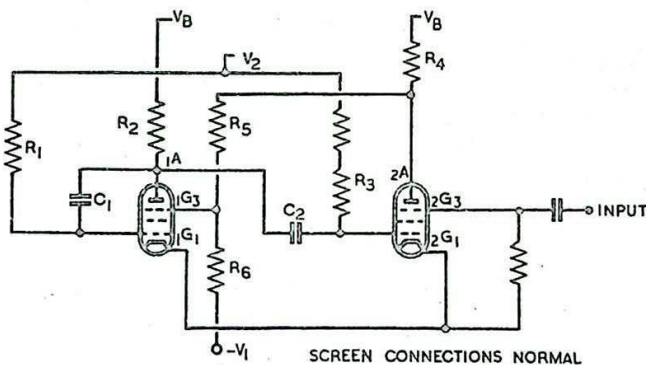


Fig. 518 - Simplified sanatron circuit.

discharges, and the voltage at $2G_1$ returns towards V_2 . As this voltage rises through cut-off, valve 2 reconducts:-

The voltage at $2A$ falls.

" " " $1G_3$ "

" " " $1A$ rises.

" " " $2G_1$ "

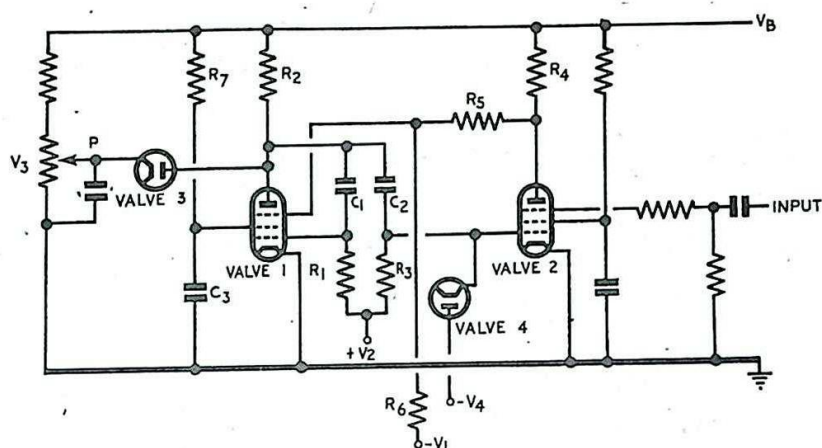
still further so that the action is cumulative.

This cumulative action ceases when the anode current in valve 1 is cut-off.

C_1 and C_2 recharge as the voltage at $1A$ returns to V_B , and the circuit reverts to its initial state.

Fig.

519 shows a practical form of the sanatron circuit, and the waveforms of the voltages at the electrodes are shown in Fig. 520.



TYPICAL VALUES OF COMPONENTS

$R_1 = 70k\Omega$	$C_1 = 100\mu F$	$V_B = 300V$
$R_2 = 500k\Omega$	$C_2 = 100\mu F$	$-V_1 = -320V$
$R_3 = 150k\Omega$	$C_3 = 0.5\mu F$	$V_2 = 200V$
$R_4 = 15k\Omega$		$-V_4 = -5V$
$R_5 = 100k\Omega$		
$R_6 = 250k\Omega$		
$R_7 = 12k\Omega$		

Fig. 519 - Practical sanatron circuit.

A diode, valve 3, is included to

limit the anode voltage of valve 1. During the recharging period the anode voltage returns towards V_B but as the diode conducts is clamped at the level V_3 , determined by the setting of the potentiometer slider P. This diminishes the recovery time of the Miller valve (interval (3)). The amplitude, and hence the duration, of the output pulse can be controlled by the potentiometer setting.

A second diode, valve 4, limits the negative excursions of the grid of valve 2 to a value $-V_4$, just sufficient to ensure that valve 2 is cut-off (about -5 volts). This reduces the interval (4) between the end of the run-down and the instant at which valve 2 reconducts.

It may be noted that the same rapid fall of voltage at $1A$ during interval (2) is communicated to the control grids of both valves. This is sufficient to cut off the current in valve 2, but not in valve 1. Hence if the valves are otherwise alike, they must be adjusted so that the cut-off voltage is lower for valve 1 than for valve 2. This may be accomplished by using a larger screen dropping resistance for valve 2.

Alternative methods of triggering, which utilise the amplification of valve 2, may be employed. The method already described, using a negative pulse applied to the suppressor grid, requires a triggering pulse amplitude of about 60V. If a negative-going pulse is applied to $2G_1$, or to $1A$ (actually to the cathode of the diode, valve 3) an amplitude of about 15 volts is sufficient. If the grid-triggering method is employed an additional diode is

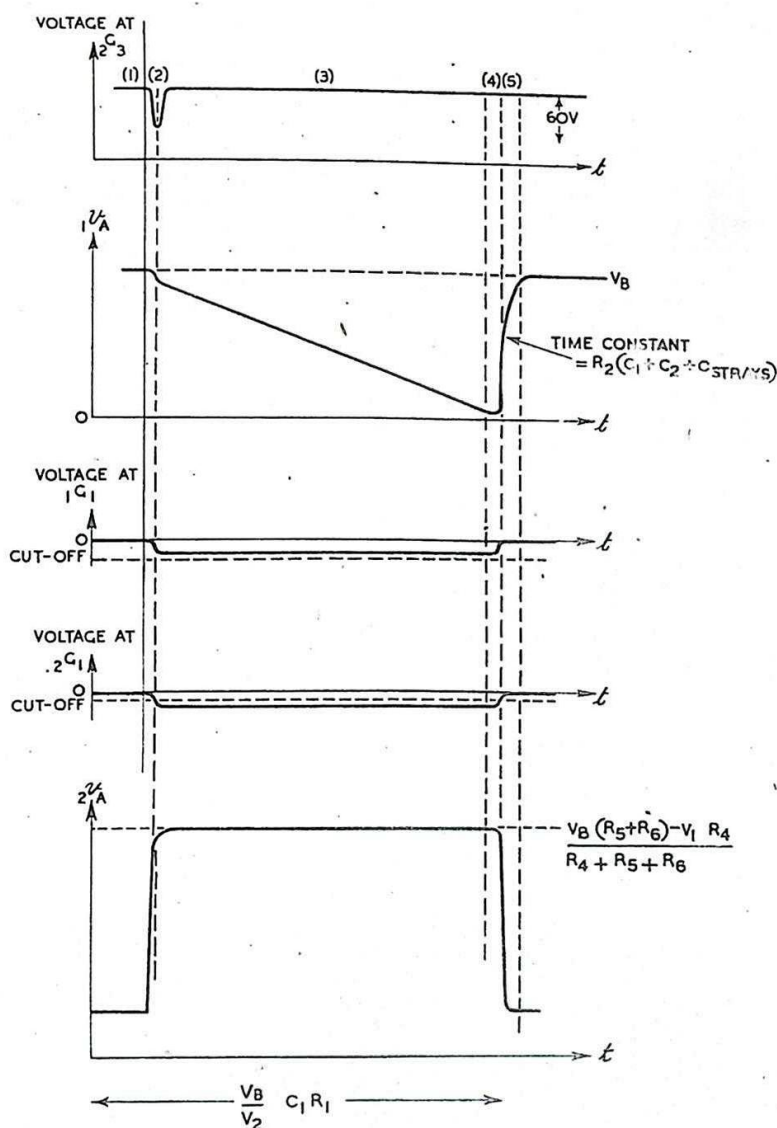


Fig. 520 - Action of sanatron circuit.

necessary, as shown in Fig. 521, to prevent the trailing edge of the triggering pulse from causing the valve to reconduct. If triggering pulses are applied to the cathode of valve 3 a cathode load resistor must be inserted, since in the circuit of Fig. 519, the cathode is decoupled to earth.

Some advantages may be derived from the use of the screen grid of valve 1 as a control electrode in place of the suppressor grid, particularly if a very fast run-down is required. The rate of run-down is limited by the anode current, which depends on the screen voltage, and high values of screen voltage lead to excessive screen dissipation. The fact that the screen is held at a low voltage except during the run-down interval reduces the screen dissipation and allows a higher screen voltage to be used, so that a faster run-down may be achieved.

The modifications which are needed to the arrangements of Fig. 519 to convert the circuit to one using screen-grid control are shown in Fig. 522.

This method of controlling the Miller valve allows the use of a beam tetrode in place of a pentode, so that larger values of anode current are available for the same screen dissipation. This further increases the permissible run-down speed.

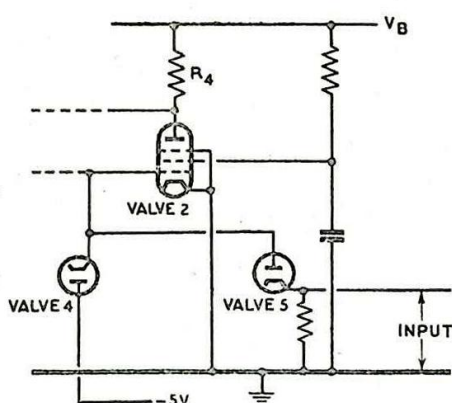
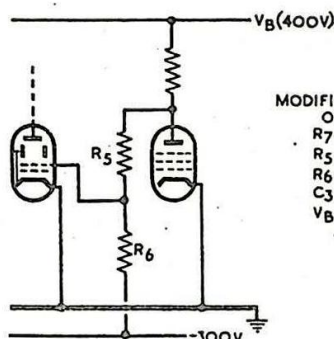


Fig. 521 - Sanatron circuit:
control grid triggering.



MODIFICATIONS TO CIRCUIT
OF FIG 2B
R7 REMOVED
R5 = 15kΩ
R6 = 60kΩ
C3 REMOVED
VB INCREASED TO 400V

Fig. 522 - Sanatron circuit:
screen grid control.

11. The Phantatron

The Phantatron is a one-valve relay, similar in action to the two-valve Sanatron circuit. A simplified diagram of a phantatron circuit is shown in Fig. 523. The anode and control grid are connected by the condenser which typifies the Miller time-base arrangement. The suppressor grid and cathode connections give rise to the cumulative action which operates the relay.

Initially, the control grid is limited at cathode potential by grid current flowing through R_1 . The cathode voltage is held at 20 or 30 volts above that of the suppressor grid by the current through R_3 , so that anode current is cut off by the suppressor grid bias. The valve current flowing through R_3 is therefore due entirely to screen and control grid currents.

A negative-going triggering pulse is applied to G_1 . The cathode voltage falls (cathode follower action).

The cathode voltage is lowered relative to that of the suppressor grid by an amount sufficient to cause anode current to flow, i.e. the suppressor grid-cathode voltage is raised above cut-off.

As anode current flows the anode voltage falls.

This fall is communicated to G_1 via the condenser C.

The cathode voltage falls still further, so that the action is cumulative.

Hence, even when the triggering pulse is removed, anode current continues to flow, and the circuit remains temporarily in its unstable state.

The anode run-down then follows, as in the Miller time-base. No appreciable change in voltage at control grid or cathode occurs until the anode voltage bottoms. The voltage at G_1 then begins to rise more steeply, and with it the voltage at the cathode. This rise in cathode voltage with respect to that of the suppressor grid reduces the anode current.

This causes the anode voltage to rise, and with it the voltage at G_1 .

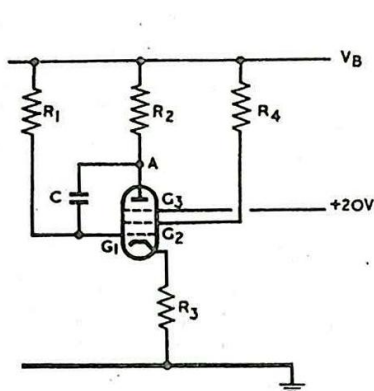
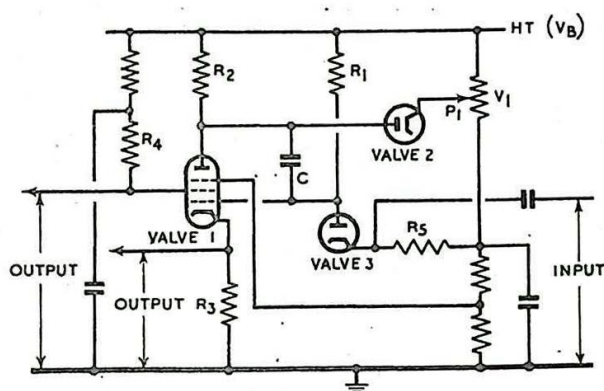


Fig. 523 - Simplified phantastron circuit.



COMMON VALUES
 $R_1 = 2M\Omega$
 $R_2 = 500k\Omega$
 $R_3 = 4k\Omega$
 $R_4 = 5.6k\Omega$
 $R_5 = 20k\Omega$
 $V_B = 300V$

Fig. 524 - Practical phantastron circuit.

The cathode voltage rises still further, so that the action is cumulative, ceasing when the anode current is cut off. C then discharges and the circuit reverts to its initial condition.

A practical phantastron circuit is given in Fig. 524, and the waveforms of the voltages at the various electrodes are shown in Fig. 525. The function of the diode in the anode circuit (valve 2) is the same as that of the corresponding diode in the sanatron circuit,

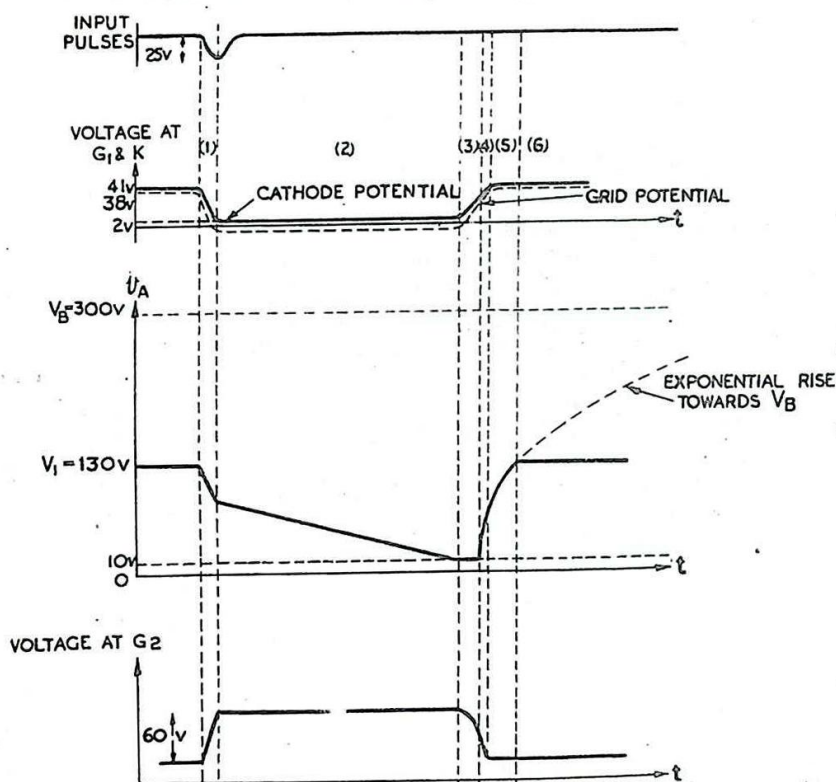


Fig. 525 - Action of phantastron circuit.

ensuring a rapid return of the anode voltage to its steady value, during intervals (4) and (5). The setting of the potentiometer P1 determines the amplitude and hence the duration of the run-down interval (2). The rate of run-down is given by:-

$$\frac{dV_a}{dt} = -\frac{V_B}{CR} \quad (\text{This assumes that } R_3 \ll R_2, \text{ as in the case described}).$$

The other diode (valve 3) isolates the triggering circuit from the control grid after the initial fall in voltage at G1, otherwise the run-down would be affected. The voltage at which its cathode is held (usually about 20V. above that of the suppressor grid of valve 1)

determines the initial and final values of the voltage at G_1 .

The anode voltage bottoms during the interval (3) and it is at the end of this interval that the anode current begins to fall and the second cumulative action is initiated. This action occupies interval (4), during which the screen grid, cathode and control grid return to their steady potentials.

As shown in Fig. 524, the screen grid is not fully decoupled, since this is the only electrode from which a positive-going pulse is obtained.

An alternative method of varying the duration of the output pulse is shown in Fig. 526. The setting of the potentiometer P_2 determines the rate of run-down and hence the duration of the output pulse.

The method of triggering described above requires, with the components given, a negative-going pulse of about 25 volts amplitude. Alternative triggering

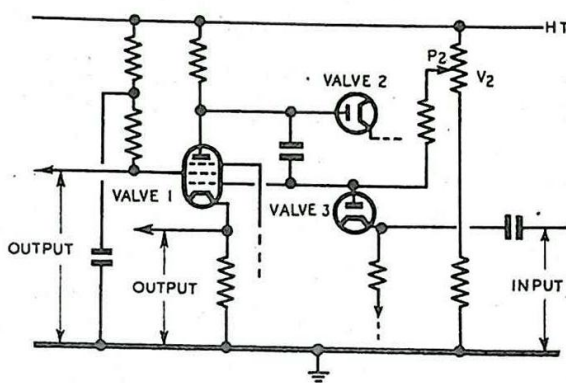


Fig. 526 - Phantastron circuit; alternative method of control.

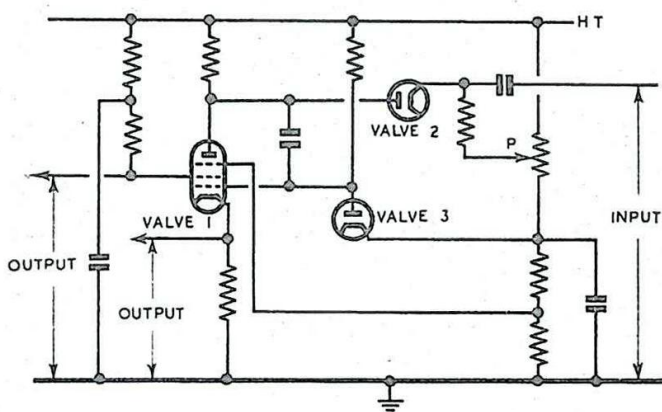


Fig. 527 - Phantastron circuit alternative triggering arrangement: negative-going pulse to μA .

arrangements are illustrated in Figs. 527 and 528. A negative-going pulse of about 40V amplitude applied to the anode (Fig. 527) or a positive-going pulse of about 25V amplitude to the suppressor grid (Fig. 528) of valve 1, is required.

The phantastron is a useful circuit in systems where the number of valves used must be kept to a minimum. It also has the advantage not possessed by the sanatron of providing both positive-going and negative-going rectangular output pulses. These can be made of approximately the same amplitude by a suitable choice of cathode and screen resistors. There are, however, several disadvantages which arise from the use of the single-valve circuit. The initial fall in voltage at the cathode (interval (1)), Fig. 526, is accompanied by a corresponding fall at the anode, so that the amplitude of the linear run-down is reduced, and a large fraction of the available supply voltage is wasted. To minimise this initial voltage

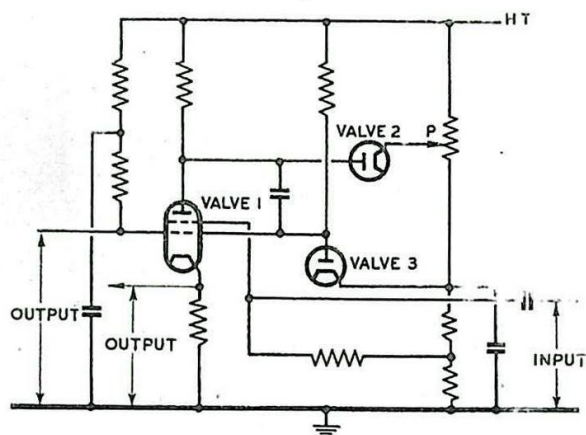


Fig. 528 - Phantastron circuit: alternative triggering arrangement: positive-going pulse.

change a valve with a very short suppressor grid base is required. Certain valves are available which have such a short base (e.g. CV1116, 3 to 15 volts), but this restricts the choice of valves so that other characteristics, which were considered in the two-valve circuit, have to be ignored.

Furthermore, the action of the phantastron is very much dependent on the choice of valves, whereas the valves used in a sanatron may be changed for others of the same type without the characteristics of the relay being appreciable affected. This is due partly to the fact that the cumulative action at the end of the linear run-down occurs over the region of control-grid cut-off in the case of the sanatron and over the region of suppressor-grid cut-off in the case of the phantastron. The latter action is not so precise as the former, and is more dependent on the particular valve used.

CHAPTER 11

TIME BASES AND TIME-BASE GENERATORS

INTRODUCTION

1. General

If it is required to draw a graph showing the change of any quantity with time, one of the first things to do is to select a suitable time scale, and mark it along one of the axes (usually the X-axis in Cartesian Co-ordinates). A time scale is needed for the measurement of range by a radar equipment, and is normally produced by deflecting the beam of a CRT so that the spot sweeps across the fluorescent screen. A similar time scale is required if any variations of voltage or current with time are to be examined.

A multiplicity of terms has arisen in connection with the provision of a time scale. Let us now consider the meanings implied by the use of such terms in subsequent sections.

The deflection of the spot of a CRT, when it is defined in relation to time, is called a Time Base. This term should be distinguished from the term Trace, which is a general term for the pattern appearing on the screen of a CRT however produced.

The term Scan is frequently used with a variety of meanings, and is correctly used in television technique where the CRT picture is produced as a result of scrutinising a light sensitive surface point by point. In radar systems the term should be reserved for use in connection with the scanning of an area by an aerial array.

A Time-Base Generator is an apparatus for producing a voltage or current varying with time which could be used to establish a time base. A time-base generator is sometimes referred to simply as a time base, but this latter term is reserved here for the actual deflection of the spot on the screen of a CRT.

The output voltage (or current) of a time-base generator is referred to as a Time-Base Voltage (or Current). Sometimes this output is used for measuring time intervals, although it is not used to deflect a CRT beam. The output, and the apparatus producing it, are still, however, qualified by the adjective time-base. The terms Sweep Voltage and Time-Base Voltage are synonymous, but the former term is retained for the sake of the graphic idea it conveys.

In order to clarify our ideas on what is meant by a time base let us consider a specific example.

If a voltage is applied between a pair of deflector plates of a CRT the spot on the screen is deflected by an amount which is proportional to the magnitude of the voltage. Suppose a time-base generator produces a voltage which increases linearly with time; (Fig. 529). If this voltage is applied between the X-plates of a CRT the spot moves across the screen in a horizontal direction, and a time base is formed. In this particular case, assuming uniform

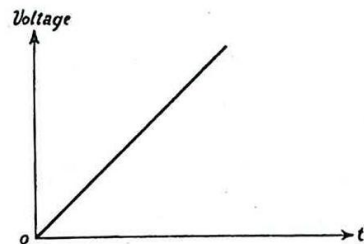


Fig. 529 - Voltage varying linearly with time.

deflection sensitivity, the movement of the spot takes place at a uniform velocity, i.e. the spot moves through equal distances in equal time intervals.

Suppose a rectangular voltage pulse is applied between the Y plates of a CRT at a time T after the instant at which the time base starts. Then the time base is disturbed and a rectangular waveform is produced on the screen of the CRT; (Fig. 530). Then the distance between the start of the time base and the leading edge of the waveform is a measure of the time T .

Similarly, the distance between the leading and trailing edges of the waveform is a measure of the duration of the applied pulse. Instead of being used to produce a deflection at right angles to the direction of the time base the rectangular voltage pulse can be employed to vary the brilliance of the spot, so that part of the time base is brighter than the remainder (Sec. 13).

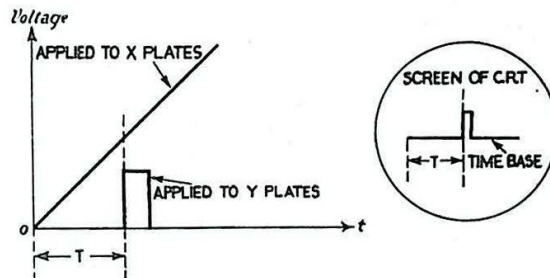


Fig. 530 - Pulse applied to Y plates.

So far it has been assumed that the spot on the CRT screen makes a single sweep. Such a sweep, known as a Single-Stroke Time Base, may be used for the examination of non-repeating voltages or currents (transients). In radar, however, a time base is required which repeats itself at regular intervals. Then if the voltage to be examined is periodic and is synchronised to the time-base voltage, the resulting picture is stationary on the screen of the CRT.

2. Linearity

The time scale on a graph is generally uniform, i.e. equal lengths on the time axis correspond to equal time intervals, and it is usually desirable that the movement of the spot forming a time base should be similarly uniform. If the graph of time base voltage against time is a straight line, as in Fig. 529, the time-base voltage is said to be linear. It should be noted that even though the time-base voltage is linear, the spot-deflection or time base will not be linear unless the CRT deflection sensitivity is uniform.

In general, the term linear implies that the quantity which is measured is proportional to time, and is not restricted to straight-line time bases. For example the time base may be circular, and time intervals measured in terms of the angular deflection of the spot. Provided this angular deflection is proportional to time the time base is linear.

3. Flyback

If the time base takes the form of a repetitive sweep along a straight line the spot on the screen of the CRT must return to its starting point after each sweep so as to be ready for the succeeding one. This return movement is called the Flyback.

The flyback is an unavoidable complication; its duration occupies uselessly a portion of the period of the waveform under

examination, and it may confuse or obscure the forward sweep by superimposing its own trace. Both troubles are mitigated by making the flyback rapid, so that the ratio of sweep time to flyback is large. Under such circumstances only a small proportion of the waveform under examination is lost. Also, provided saturation of the CRT screen does not occur, the brilliance of the flyback is much less than that of the time base itself because the spot moves much more quickly during flyback than during the forward sweep.

For the examination of recurrent voltage variations in which the whole cycle is of interest a time-base voltage of the type shown in Fig. 531 is suitable; the duration of the flyback is negligible compared with that of the forward sweep.

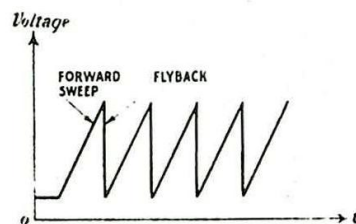


Fig. 531 - Continuous saw-tooth voltage.

In radar we are often concerned with only a small fraction of the whole cycle of the time-base voltage, so that it is not essential for the flyback to occur in the shortest possible time. It is still, however, necessary to avoid confusion

between the sweep and the flyback. This is frequently achieved by reducing the voltage at the CRT grid during the flyback interval, the brilliance control being so adjusted that the forward sweep is visible and the flyback invisible. Alternatively, the grid potential may be increased during the forward sweep to give increased brightness. This "blacking-out" or "brightening" is accomplished by applying rectangular pulses of voltage at appropriate instants to the grid or cathode of the CRT (see Chap. 6).

Such pulses can be readily obtained as a by-product of the time-base generator,

especially if such a generator includes a relay or relaxation oscillator. Alternatively, the sweep voltage can be applied to a C-R circuit the time constant of which is short compared with the duration of the sweep voltage. In this case the voltage developed across the resistor is in the form of approximately rectangular pulses suitable for making the flyback invisible; Fig. 532. (see Chap. 2 Sec. 15).

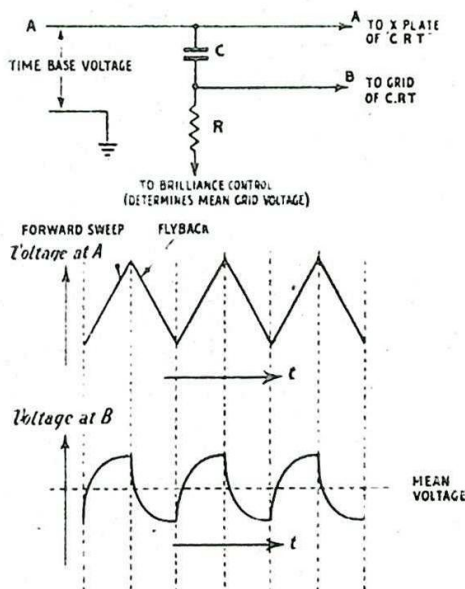


Fig. 532 - Production of pulses for brightening forward sweep and blacking out flyback.

4. Time Bases Derived from a Source of Steady Voltage

The time bases discussed in the previous section are derived in the main from an alternating supply. It is, however, far more common to obtain the time base voltage (usually of sawtooth form) from a source of steady voltage. Time bases derived in this way are of the straight line type, and the usual aim is to make them linear.

Time-base generators of this sort can be considered under two main headings (i) capacitive (ii) inductive. In type (i) the time-base voltage is derived from the change of voltage across a capacitance as it charges or discharges through a resistance. In type (ii) the change of current through an inductance and a resistance in series is used. Type (i) has much the wider application in radar equipments.

In both types a switching device is necessary to change the conditions of the time-base generator so that both a forward sweep and flyback are formed. The switching device may be self-operating or externally controlled. Generally, in radar equipment the switching is controlled externally so that its timing is related to other operations. In cathode-ray oscilloscopes the switching is normally self-operating, but usually arrangements are made so that the time-base voltage can be synchronised with the voltage under examination if required.

TIME-BASE GENERATORS FOR ELECTRIC DEFLECTION

5. Fundamental Circuit of Capacitive Time-Base Generator

The fundamental circuit of a capacitive time-base generator is shown in Fig. 533. The condenser C is charged through some form of charging device, and is discharged by closing the switch S . To ensure rapid action S is an electronic switch, using either "soft" or "hard" valves.

Alternatively, the condenser may be charged during the flyback period, the time-base voltage being provided by the discharge. This latter arrangement is particularly suitable for the production of very fast time bases. The necessarily small time-base resistor is then used in the discharge circuit only, and the charging resistor can be made relatively large. This arrangement reduces the likelihood of damage to the supply circuits by a break-down of the insulation across the time-base condenser.

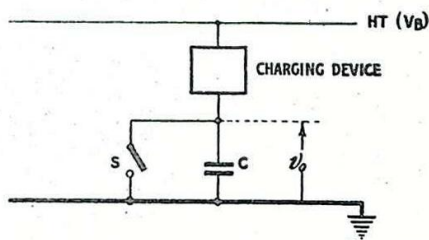


Fig. 533 - Fundamental circuit of capacitive time-base generator.

In the circuit of Fig. 533 the condenser charges while the switch is open, and the voltage V_0 across it rises. This voltage is proportional at any instant to the charge on the condenser. If this charge increases linearly with time, i.e. if the current flowing through the charging device into the condenser is constant, V_0 increases linearly with time. In other words, if the time base derived from this generator is to be linear the charging device must pass a constant current.

When the switch is closed the condenser discharges. This discharge can be considered as instantaneous provided the resistance of the discharge circuit is small.

If the discharge circuit is of the self-operating type, the switch closes when the voltage across the condenser rises to some critical value V_S , and opens when it falls to some lower critical value V_E . The sawtooth waveform of the voltage across the condenser is shown in Fig. 534. The rate of rise of voltage can be altered by adjustment of the charging device so that the value of the charging current is

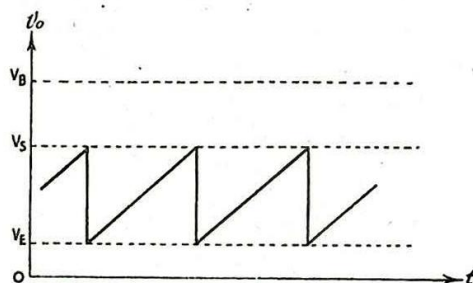


Fig. 534 - Action of circuit of fig. 533.

changed. The rate of rise of voltage determines the velocity at which the spot traces out the corresponding time base on the screen of the CRT. Hence, the control which alters the charging current is usually named the Velocity Control. Provided V_S and V_E are constant, alteration of the velocity control does not change the amplitude of the sawtooth voltage, i.e. does not change the length of the time base. It follows that the setting of this control determines the time base repetition frequency; (Fig. 535). Alteration of the voltage differences $V_S - V_E$ alters the amplitude of the sawtooth voltage, i.e. changes the length of the time base. A control which performs this function is termed the Amplitude Control. The amplitude control also alters the repetition frequency of the sawtooth voltage; (Fig. 536).

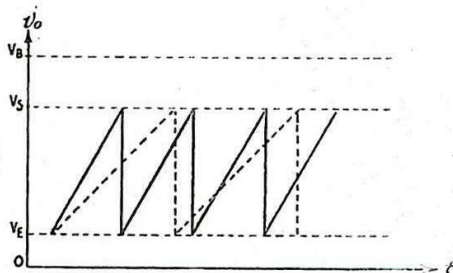


Fig. 535 - Case of alteration of velocity control changing repetition frequency.

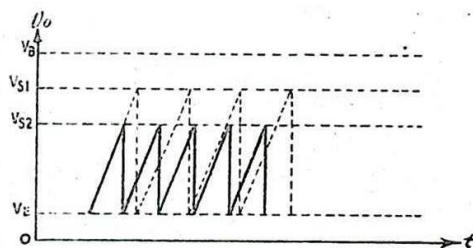


Fig. 536 - Case of alteration of amplitude control changing amplitude and repetition frequency.

If the discharge circuit is of the externally controlled type, the duration T of the rise of voltage across the condenser is fixed by the external control. Alteration of the charging current again changes the rate of rise of voltage, i.e. the velocity of the spot. Since the duration of the time base is fixed, adjustment of the velocity control is the only means of adjusting the amplitude of the time base voltage; (Fig. 537).

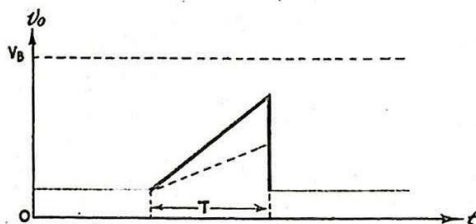


Fig. 537 - Case of alteration of velocity control changing the amplitude.

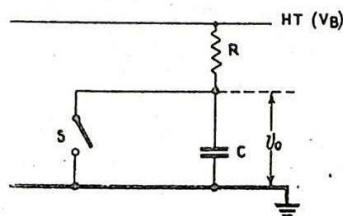


Fig. 538 - C-R charging circuit.

6. The Charging Circuit of a Capacitive Time-Base Generator

The simplest charging device is a resistor R , and if it is used the circuit of Fig. 533 becomes that shown in Fig. 538. If the switch is open, the voltage across the condenser rises exponentially towards the value V_B . Provided only a small fraction of the total rise of voltage is used to form the time base, the latter can be considered to be linear;

(Fig. 539). However, a disadvantage of this method of forming a linear time base is that either V_B must be very large or else amplifiers must be used if a deflection voltage sufficient to obtain a time base of reasonable length is to be available.

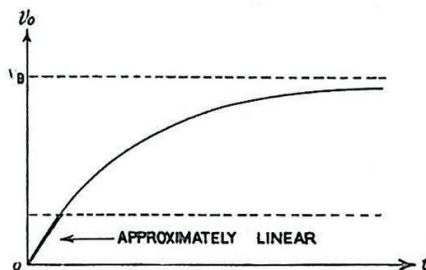


Fig. 539 - Production of approximately linear time-base voltage.

Control of velocity is obtained by adjustment of the value of the resistor, since such an adjustment alters the rate of rise of voltage, i.e. the velocity of the spot. Alternatively the capacitance of the condenser can be altered.

If the circuit is operating between fixed voltage limits, the degree of linearity is the same for all values of the time-constant. If, however, the duration is fixed the time-base voltage becomes more linear as the time-constant is increased.

7. Methods of Improving Linearity of Charging

Although the question of linearity is often mentioned in connection with time bases it is not necessarily of overmastering importance. In waveform examination, lack of time-base linearity will of course lead to a distorted picture. For example, Fig. 540 shows the waveform of a sinusoidal voltage displayed on an exponential time base. However, for ranging purposes importance of linearity depends to some extent on the method adopted for transmitting range information. Where the operator reads ranges from a scale mounted on the face of the tube, error can be avoided by making the scale correspondingly non-linear. If, however, range is measured by mechanical means (e.g. by turning a handwheel to keep a crosswire or marker in coincidence with an echo) and conveyed by a transmission system which must have a linear calibration, construction of the range measurement system is considerably simplified if the time base is made linear. Further, the time base must be linear if uniform target discrimination is to be provided at all points along the time base.

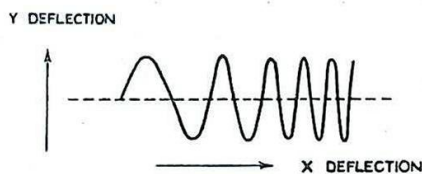


Fig. 540 - Waveform of sinusoidal voltage using exponential time-base.

When a high degree of linearity has been obtained at the time-base generator care must be taken to prevent distortion of the time-base voltage between this point and the deflector plates.

The time-base voltage is usually applied either directly or after amplification to the plates of a CRT, normally via a coupling condenser and leak resistor. The time-constant of this coupling circuit must be long compared with the duration of the time-base voltage otherwise distortion is introduced (see Chap. 2 Sec.15). In any case even if a constant-current charging device is used, the change of voltage across the time-base condenser is not absolutely linear in the presence of the coupling circuit, since the constant current flows into a network which is no longer purely capacitive.

The capacitance across the deflector plates is usually considerable, and when deflector plate current flows the conductance between the plates becomes appreciable, so that the input impedance of the deflection system is not as large as would be desirable. For this reason in order to ensure that the full time-base voltage is developed between the plates at all times the output impedance of the time-base generators must be small. This may be ensured by inserting a power amplifier, such as a cathode follower, between the time-base condenser and the deflector plates.

Various methods of improving the linearity of the time-base voltage derived from a capacitive generator are given below in paragraphs (i) to (iv). The Miller Time-Base Generator, which provides one of the most accurate methods of achieving linearity, is described separately in Sec. 10.

- (i) If only a small fraction of the total possible voltage change across the condenser is used the resulting time base is nearly linear. This method has already been discussed.
- (ii) A constant current device can be substituted for the charging resistor. Such a device is a pentode, operated above the knee of its $I_a - V_a$ characteristic.

The $I_a - V_a$ characteristics of a typical R.F. pentode, operating at a fixed screen potential for various values of grid bias, are shown in Fig. 541. From these characteristics it is apparent that if the grid bias is, say $-4V$, the anode current remains sensibly constant for anode voltages greater than about 50V. If such a valve is incorporated in the circuit of Fig. 542, when the switch is opened the condenser charges almost linearly with time until the anode voltage falls to 50V. For example, if the supply voltage is 300V, the voltage across the condenser, i.e. the time-base voltage, can have a magnitude as much as 250V without there being any appreciable departure from linearity.

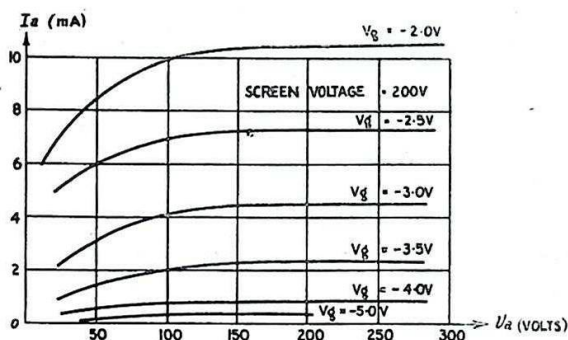


Fig. 541 - $I_a - V_a$ characteristic of typical HF pentode.

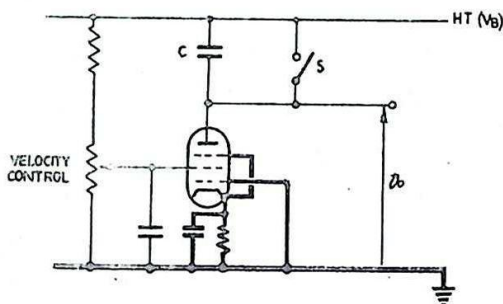


Fig. 542 - Charging circuit employing HF pentode.

The screen potential can be adjusted to provide velocity control, i.e. to change the charging current, while the grid is held at a fixed potential. In an alternative arrangement the condenser is in the cathode circuit of the valve, the time base voltage being taken from the cathode. In this case the screen must be decoupled to cathode, since variations in screen-cathode voltage would cause a change in charging current. The grid is connected to a point a few volts below cathode potential using a small cathode bias resistor.

The current through the pentode is not absolutely constant; for it to be so the slope resistance of the valve would have to be infinitely large. The greater this slope resistance the closer is the approach to linearity of the voltage across the condenser. Any method of increasing the effective value of the slope resistance leads to an improvement in linearity.

The circuit arrangements shown in Figs. 543(a) and (b) are useful in achieving this object. Fig. 543(a) shows a simple current feedback arrangement which, as described in Chap. 7, Sec. 16, increases the effective slope resistance of the valve. Fig. 543(b) shows a similar arrangement in which a second valve is used, resulting in a somewhat better current stabilisation.

(iii) If amplification of the time-base voltage is necessary, it is possible to arrange circuit conditions so that the amplifier introduces distortion which compensates for the non-linearity of the exponential time-base voltage. Exact compensation is unlikely.

For reasons given in Chap. 6 it is advisable to use balanced circuits to feed the CRT deflector plates. Fig. 544 shows how a phase-inverting amplifier employed in a paraphase circuit may be used to provide the right kind of distortion to counteract the non-linearity of the initiating time-base voltage. The amplifier valve is so biased so that it is operating on the lower bend of its dynamic characteristic.

In the circuit shown at (a), C_1 and C_2 , which together form the time-base capacitance, act also as a potential divider, so that the amplitude of the amplifier output voltage v_2 applied to the X_2 plate is equal to that of the voltage v_1 applied to the X_1 plate. The conditions are normally arranged so that when the fraction

$\frac{C_1}{C_1 + C_2}$ of the voltage v_1 (Fig. 544(b)) is fed to the amplifier,

the output v_2 is curved as shown at (c). The deflecting voltage

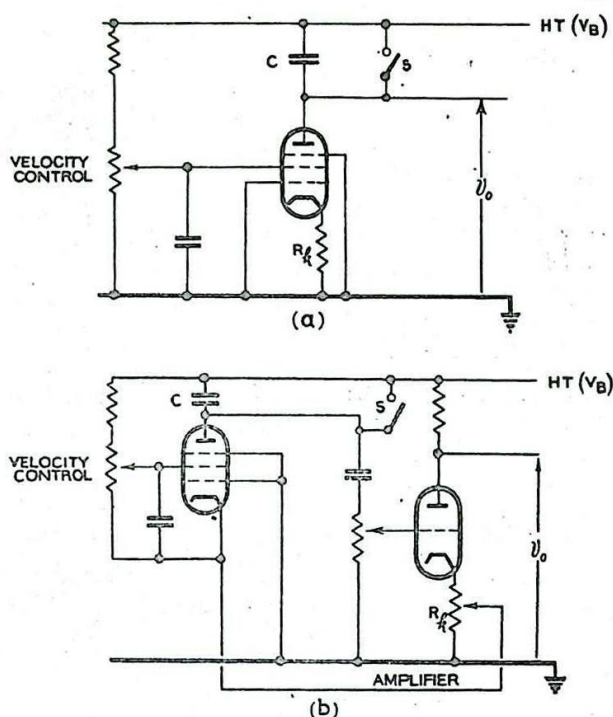


Fig. 543 - Circuits incorporating feedback to improve linearity of charging.

$(v_1 - v_2)$ is then approximately linear as shown at (d).

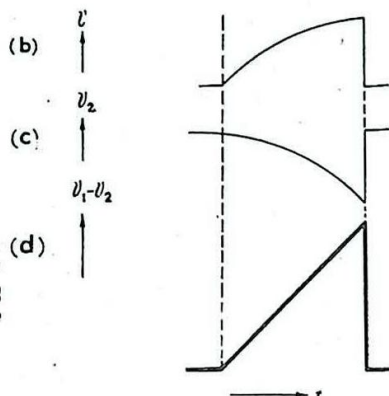
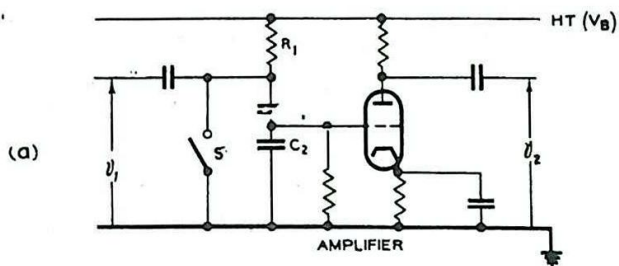


Fig. 544 - Action of circuit arrangement to counteract effect of non-linearity of time-base voltage.

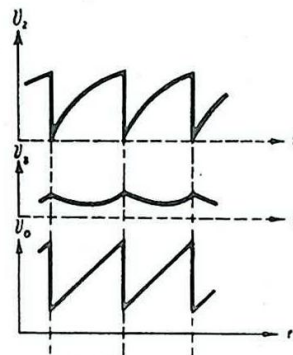
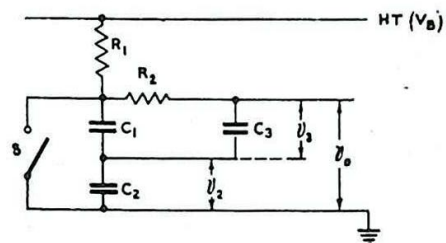


Fig. 545 - Action of additional C-R to improve linearity.

The circuit described above has the advantage that no charging pentode is required. However the adjustments to the circuit are critical. If, for example, the change of condenser voltage takes place between fixed time limits (externally controlled), any alteration in the time-constant of the circuit leads to an alteration in the curvature of the fraction of the voltage v_1 which is applied to the amplifier.

(iv) The output of a simple C - R time-base circuit can be made more linear by the addition of another C - R network. One form of such an arrangement, suitable for a self-operating time-base generator is shown in Fig. 545.

Here the charging resistance is R_1 and the main time-base capacitance consists of C_1 and C_2 in series.

Assume that the condensers, C_1 , C_2 and C_3 are charged. When the switch S is closed, C_1 and C_2 discharge rapidly. C_3 also discharges, but does so relatively slowly since the time-constant $C_3 R_2$ is long.

The switch opens when C_1 and C_2 are discharged, and at this instant the voltage across C_3 is still comparatively large. C_1 and C_2 commence to charge, but since the voltage across C_3 is greater than that across C_1 , C_3 still discharges. This discharge of C_3 continues until the voltages across C_1 and C_3 become equal, and then C_3 is charged from C_1 . The voltage rise across C_2 is approximately

exponential, whilst as we have just explained the voltage across C_3 falls at first, but after a time starts to rise.

As shown in Figs. 545(b) and (c) the curvatures of the voltage variations v_2 and v_3 are of opposite sign. Hence, if suitable values of components are chosen, the output voltage, which is the sum of these two voltages, can be made substantially linear.

8. Discharge Circuits

(i) Switching by means of Soft Valves

A gas-filled diode (neon valve) was the earliest form of switching valve used in a time-base circuit. The operation of this type of discharge valve has been described in Chap. 10 Sec. 1. Owing to the limitations set by the fixed striking voltage and extinction current neon time-base circuits are practically obsolete, giving place in the first instance to circuits using gas-filled triodes. A typical arrangement of a gas-filled triode circuit is shown in Fig. 546. If the grid bias is suitably adjusted the circuit operates as a free-running relaxation oscillator, the operation being identical with that of the neon circuit mentioned above. By variation of the grid bias the striking voltage and hence the amplitude of the time base voltage can be controlled. The action can be synchronised by the injection of a controlling voltage between grid and cathode.

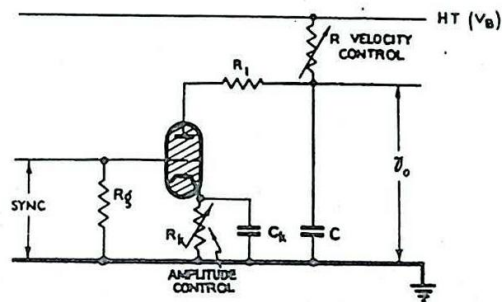


Fig. 546 - Switching by means of gas triode.

The principal advantages of using a gas-filled triode as a switching valve are the low resistance presented by the valve during the discharge, and the ease with which the circuit can be made free-running. The choice of component values, however, is limited by the uncontrollable value of the "extinction current". Ultimately the use of gas-filled discharge valves is restricted by the comparatively long de-ionisation time, of the order of a millisecond. It is not normally practicable to use soft discharge valves if the time-base frequency is more than about 40 Kc/s.

(ii) Switching by means of Hard Valves

The place of the thyatron in the discharge circuit of Fig. 546 can be taken by a hard valve, provided means are available by which the valve current can be switched rapidly on and off. With the hard valve there is no question of ionisation and de-ionisation of gas. In general, a time base generator using hard-valve switching is more definite in its operation and capable of operating

at a higher repetition frequency than one employing a soft valve.

The basic circuit in which hard-valve switching is employed is shown in Fig. 547. If the valve is conducting heavily the time-base condenser discharges rapidly, whilst if the anode current is cut off the condenser charges through a resistor. In ranging equipment the valve is usually a pentode, and switching is accomplished by alteration of the voltage at one of the grids.

By variation of the mark-to-space ratio of the input switching voltage the output voltage can be given various forms such as those illustrated in Figs. 548(a) and (b).

If the charging circuit, together with the switching valve (or valves), is incorporated in a relaxation oscillator, the circuit as a whole can be considered as a self-operating time-base generator. An example (Fleming-Williams circuit) has already been discussed in Chap. 10, Sec. 6. In radar the pulse

which operates the switching valve of Fig. 547 is normally derived from a separate relaxation oscillator or relay. Since in this case the charging or discharging of the condenser does not control the operation of the switching valve, the circuit should be considered as a time-base generator which is externally controlled. When a high voltage supply circuit is used for a time-base generator the discharge valve must be chosen to withstand the maximum voltage without appreciable leakage current. A pentode with a top-cap is normally employed. Such a valve is capable of withstanding as much as 4000 volts between anode and cathode without causing a leakage current of more than a few microamps.

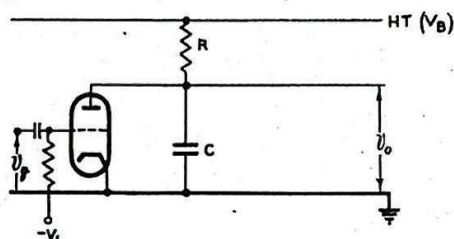


Fig. 547 - Hard valve switching circuit.

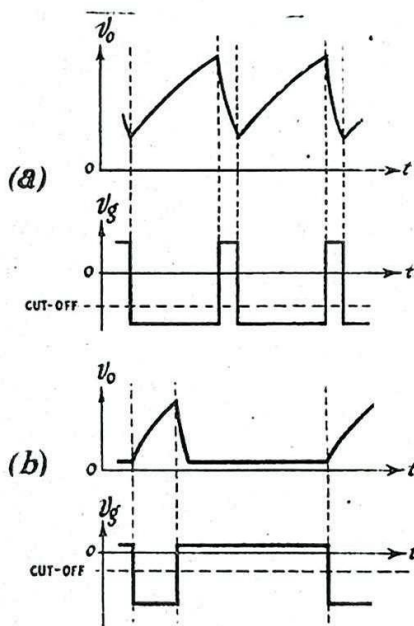


Fig. 548 - Action of hard valve circuit of fig. 547.

9. Puckle Time-Base Generator

This time-base generator, the circuit of which is shown in Fig. 549, is of the self-operating type. Valves 1 and 2 together form a relaxation oscillator. Valve 3 is a pentode, the "constant" current through which is used to charge the time-base condenser C_3 . Fig. 550 shows the relevant waveforms. An essential characteristic of this circuit is the direct connection between $1A$ and $2G$. The connections to suppressor and control grids of valve 1 are reversed in some variants of this circuit.

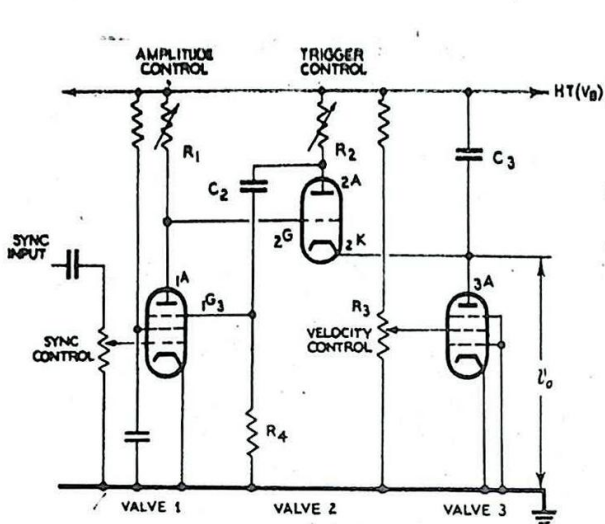


Fig. 549 - Puckle time-base generator.

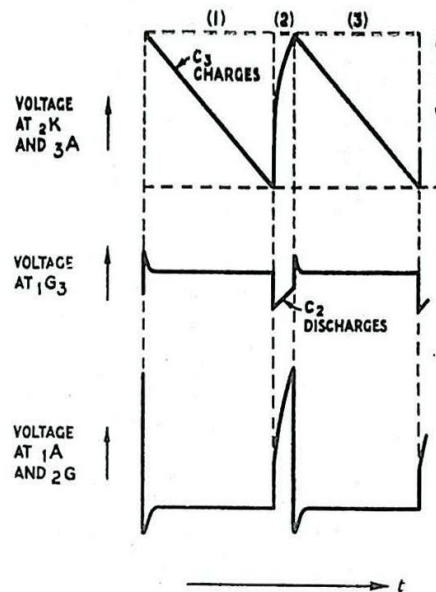


Fig. 550 - Action of Puckle time-base generator

Interval (1)

During the interval (1) C_3 charges through valve 3. Valve 2 is cut off, the voltage at $2G$ ($1A$) being at a low level, since valve 1 is conducting. As the voltage at $2K$ approaches that at $2G$, valve 2 begins to conduct.

Current flows in valve 2.

The voltage at $2A$ falls.

The voltage at $1G_3$ falls.

The voltage at $1A$ rises, and with it the voltage at $2G$. This further increases the current in valve 2 so that the action is cumulative. This action ceases when the anode current of valve 1 is cut off and valve 2 is conducting heavily.

Interval (2)

C_3 discharges through valve 2 and R_2 in series, as

3A (2K) rises towards V_B . C_2 discharges as $1G_3$ returns towards earth potential. A point is reached (suppressor grid cut-off) at which anode current begins to flow again in valve 1.

The voltage at 1A (2G) falls.

The current in valve 2 decreases.

The voltage at 2A rises and with it the voltage at $1G_3$.

The action is thus cumulative, ceasing when the current in valve 2 is cut off and valve 1 is conducting heavily.

Interval (3)

C_3 charges as the voltage at 3A returns towards earth potential, and the cycle repeats.

Duration of intervals.

The amplitude V of the voltage variation at 3A (the time-base voltage) depends on that at $2G$ (1A) since valve 2 acts as a cathode follower. This amplitude is therefore determined by the value of R_1 . The duration of interval (1) is the time taken for the voltage across C_3 to change by an amount V at a charging rate determined by the current through valve 3.

The flyback interval (2) depends on the rate of rise of voltage at $2G$, the time-constant being approximately R_1 times the stray capacitance between $2G$ and earth. This time-constant actually governs the rate of rise of voltage at $2G$. The rate of rise at $2K$ is further affected by the time-constant $C_3 (R_2 + R_V)$ where R_V is the resistance of valve 2 (about 2000Ω). Which ever of these time-constants is the larger effectively determines the rate of change of current in valve 2, and therefore the rate of rise of voltage at 2A.

This rise of voltage is transferred to $1G_3$ and thus increases the rise due to the discharge of C_2 through R_4 . The duration of the interval (2) is thus dependent upon several time-constants, but usually it is $C_2 (R_2 + R_V)$ which is decisive.

The maximum repetition frequency is ultimately limited by the flyback interval (2). The upper limit in practical circuits is about 1 Mc/s.

Controls

R_1 determines the amplitude of the time-base voltage. As indicated above, it also has an appreciable effect on the duration of interval (1) and therefore the repetition frequency.

The value of the "constant" current in valve 3 is determined by the screen voltage, so that R_3 is the velocity control. R_2 determines both the time-constant $C_2 (R_2 + R_V)$ and the amplitude of the voltage variations at 2A and $1G_3$. It therefore controls the duration of the flyback. It is normally known as the trigger control, and may be used as a vernier adjustment to the repetition frequency.

The amplitude of the voltage used for synchronisation may be varied by means of the potentiometer marked Sync. Control. It should be noted that when the repetition frequency is fixed by the synchronis-

ing signal alteration of the velocity control varies the amplitude of the time-base voltage.

10. Miller Time-Base Generator

The principle of the Miller Time-Base generator provides one of the simplest and most effective methods of achieving time-base linearity. The basis of this principle is illustrated in Fig. 551.

For changes in v_o , the output voltage of the high-gain amplifier, the corresponding changes in v_g are small. If a constant voltage V_1 much greater than v_g is applied to the input, the voltage across R is approximately constant, so that to a first approximation the condenser C is connected to a constant-current source. The condenser voltage and therefore the output voltage vary almost linearly with time.

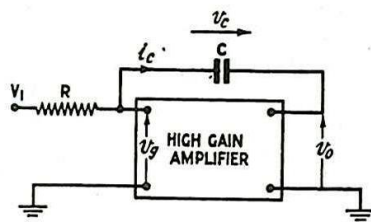


Fig. 551 - Diagram illustrating principle of Miller time-base generator.

The more detailed operation of the circuit may be described as follows. If we neglect the input admittance of the amplifier, the same current i flows through R and C .

The voltage across R is:- $V_1 - v_g$

so that the current through R is

$$\frac{V_1 - v_g}{R},$$

This is the current through C , so that the variation of v_c is given by the equation:-

$$\frac{dv_c}{dt} = - \frac{i_c}{C} = - \frac{1}{CR} (V_1 - v_g).$$

But $v_c = v_o - v_g$

Hence $\frac{d}{dt} (v_o - v_g) = - \frac{1}{CR} (V_1 - v_g) \dots\dots\dots(1)$

Provided v_g is negligible compared with V_1 and the gain of the amplifier is large, so that changes in v_g are small compared with changes in v_o , equation (1) becomes approximately

$$\frac{d}{dt} (v_o) = - \frac{V_1}{CR}$$

i.e. the output voltage (the time-base voltage) varies linearly with time.

Although any high-gain amplifier will suffice to operate this circuit, the only system in common use employs a single high-gain pentode. The arrangement is shown in Fig. 552(a). Normally the anode current is cut off by the suppressor grid being held at a low negative potential so that the anode potential is the same as that of the supply. When a positive-going pulse is applied to the suppressor

grid, raising the voltage there to zero, the anode voltage falls sharply as anode current flows. The grid voltage falls from zero, where it was limited by grid current, to a few volts negative. The exact extent of this initial fall is discussed in detail later. The anode voltage then commences the linear Run-down as C charges through R (Fig. 552(c)). When the switching pulse is removed the anode current is cut off and the circuit reverts to its initial condition as C charges. The time-constant of the rise of anode voltage is approximately CR_e . (This neglects stray capacitance in comparison with C, and the grid-cathode resistance when grid current flows in comparison with R_e).

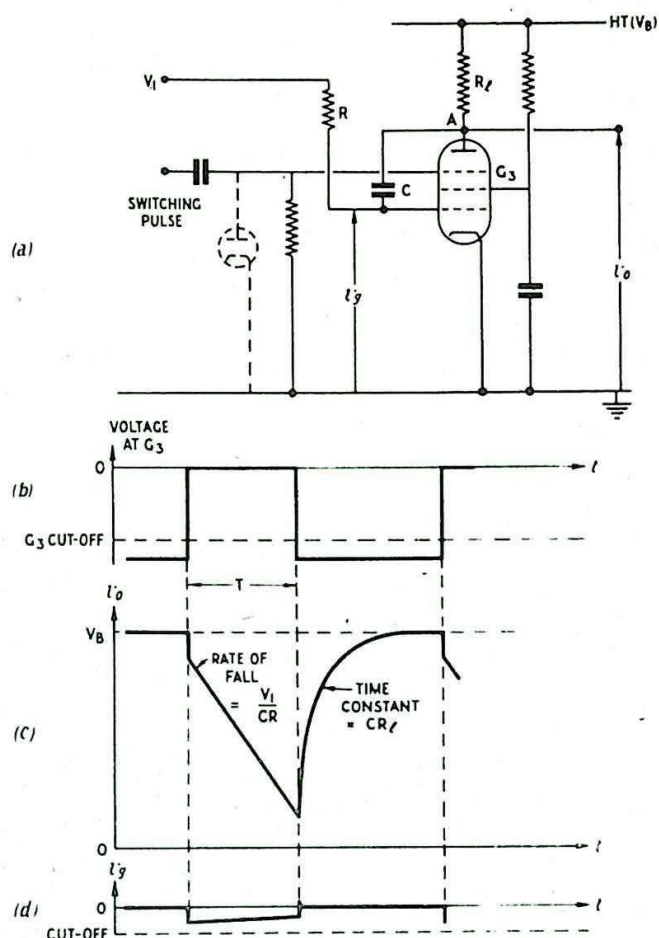


Fig. 552 - Action of Miller time-base generator.

The diode in Fig. 552(a) is a clamping diode to limit the suppressor grid voltage at earth potential.

If T is the duration of the applied pulse, the extent of the fall of the anode voltage is given by

$$T \frac{d v_0}{dt} \approx \frac{T V_1}{CR}$$

Provided this is less than V_B , the amplitude can be varied by changing R. If $\frac{TV_1}{CR} > V_B$, the anode

voltage is limited by the upper bend curvature of the dynamic characteristic. As shown in Chap. 9, Sec. 4(iv), this may occur for values of grid voltage below zero, so that subsequent changes in the voltage at the grid have little effect on that at the anode. When this occurs, as shown in Fig. 553, the time-base voltage is said to

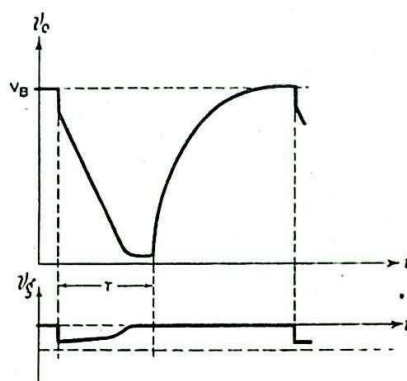


Fig. 553 - Action of Miller time-base generator when anode voltage "bottoms".

"bottom". The anode voltage remains constant so that the change in the voltage across the condenser appears as an increase in the rate of rise of grid voltage.

An alternative explanation of the Miller time-base circuit may be derived from the normal consideration of the Miller effect (see Chap. 7, Sec. 7). In the usual amplifier circuit, the input capacitance C_i of the valve is the sum of the grid-cathode capacitance C_{gk} and a feed-back capacitance $(|m| + 1) C_{ga}$, where $|m|$ is the voltage amplification, so that

$$C_i = C_{gk} + (|m| + 1) C_{ga}$$

In the Miller time-base generator the external capacitance C in parallel with C_{ga} is much greater than the interelectrode capacitances, so that the effective input capacitance is given approximately by $C_i \approx (|m| + 1) C$.

The rise in grid voltage is then similar to that of the equivalent circuit shown in Fig. 554, the time-constant being $(|m| + 1) CR$. Since the duration T of the applied voltage is of the order CR , and $|m|$ is large,

$$T \ll (|m| + 1) CR$$

so that only a small portion of the exponential voltage variation at the grid is utilised. Hence the rise in grid voltage and, consequently, the fall in anode voltage are approximately linear.

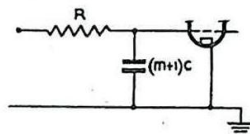


Fig. 554 - Equivalent input circuit for Miller time-base generator.

Effect of the time-base capacitance on the performance of the valve as an amplifier

So far the amplifier has been assumed to operate in a normal manner; in practice the presence of the comparatively large capacitance between anode and control grid may appreciably modify the amplifier characteristics. If the current i_c through the condenser (Fig. 555(a)) is comparable with the anode current i_a , the normal $I_a - V_a$ characteristics, superimposed on the load line, do not give the true operating conditions.

Fig. 555(b) shows how the valve characteristics can be modified to give the true conditions. For each value of v_g , the current i_c through the condenser may be determined from the relation

$$i_c = \frac{V_1 - v_g}{R}$$

If this amount is subtracted from the $I_a - V_a$

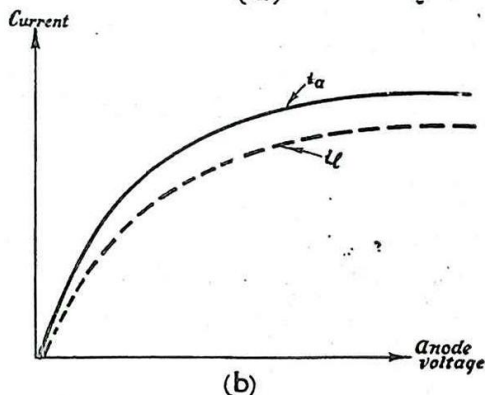
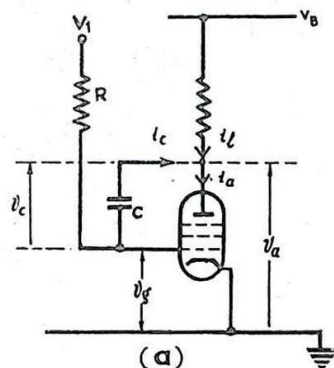


Fig. 555 - Modification to $I_a - V_a$ characteristic to give true operating conditions of Miller time-base valve.

characteristic a new curve, showing the variation of i_g with v_a , is obtained, and the intersections of the load line with a set of these curves give the true operating conditions. In most Miller time-base circuits the modification required is small since $i_g \ll i_a$, but if R is small enough to be comparable with the anode load, the condenser current is by no means negligible.

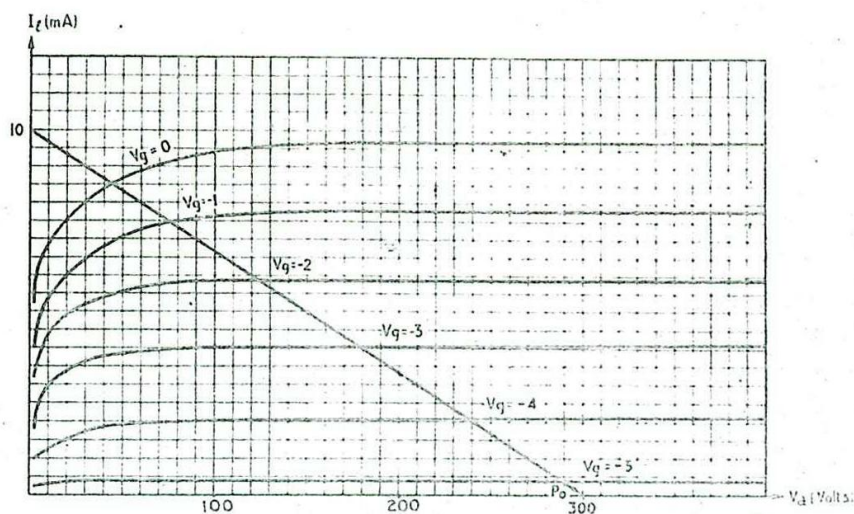


Fig. 556 - Pentode characteristics.

Fig. 556 shows a set of these modified characteristics with the load line superimposed. Initially the valve is not operating, the anode-current being cut-off by suppressor grid action, and the voltage across the condenser is approximately V_b (300V). As soon as anode current flows, the anode and grid voltages fall so that the working point P_0 is reached where $v_a = v_g = v_c = 300V$. In the case illustrated this occurs for $v_a = 295V$ and $v_g = -5V$. The anode run-down then follows the load line until terminated either by the anode voltage bottoming (due to grid current) or by the return of the suppressor grid to its cut-off value.

For any working point P , the values of v_g and v_a may be obtained from the load line.

Hence $v_g = v_a - v_g$ may be obtained.

$$\text{Since } \frac{dv_g}{dt} = -\frac{i_g}{C} = -\frac{V_1 - V_k}{CR}, \frac{dt}{dv_g} = -\frac{CR}{V_1 - v_g},$$

so that $\frac{dt}{dv_g}$ may be plotted against v_g . The area enclosed by this graph gives the time t . It is thus possible to obtain the values of v_g , and therefore of v_a and v_g , for each instant t .

An analysis of this nature reveals that for a pentode with an amplification factor of the order of 100, departure of the anode voltage from linearity is normally less than 1%, and, by a suitable choice of components, can be made considerably less.

It may be shown that the ideal shape for the dynamic characteristic is the logarithmic form shown in Fig. 557. For a given pentode the anode load, charging resistor R , and input voltage V_1 may be chosen so that the dynamic characteristic has the optimum shape. Under these

circumstances the departure from linearity may be reduced to as little as 0.01%. Such accuracy is not, however, independent of the valve used, and if an interchange of valves is necessitated the accuracy is reduced to about one part in a thousand.

If very slow time bases are required a large capacitance C must be used. If this is greater than a few microfarads the leakage resistance will be appreciable and the linearity of the output voltage will be affected. (This is not likely to arise in time-base circuits, but is important where the Miller principle is utilised in an integrating circuit using a time-constant of several seconds, such as might be required in a mechanical control system (see Chap. 18, Sec. 21).

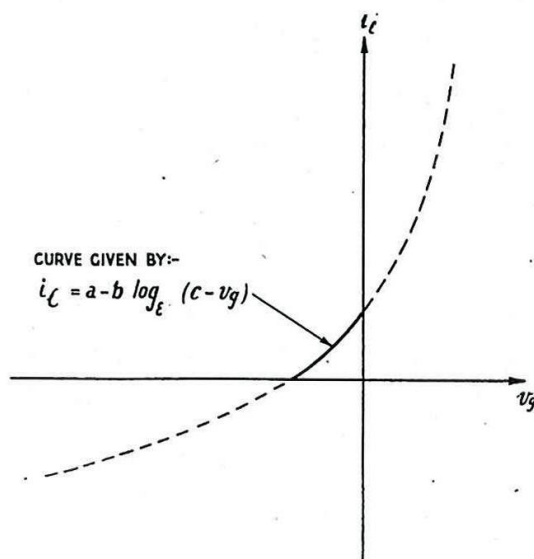


Fig. 557 - Ideal shape of dynamic characteristic.

11. Time-base Generators for Magnetic Deflection

The deflection of the spot of a CRT can be produced by changing the current in a deflector coil. If a linear time base is required, the current through the coil must rise linearly with time.

The fundamental circuit consists of a coil and resistor in series. (The coil may be the deflector coil itself, or may be inductively coupled to the deflector coil. The resistance may consist wholly or partly of the resistance of the coil).

Consider the effect of a steady voltage instantaneously applied to such a circuit. The current in the circuit grows according to the equation

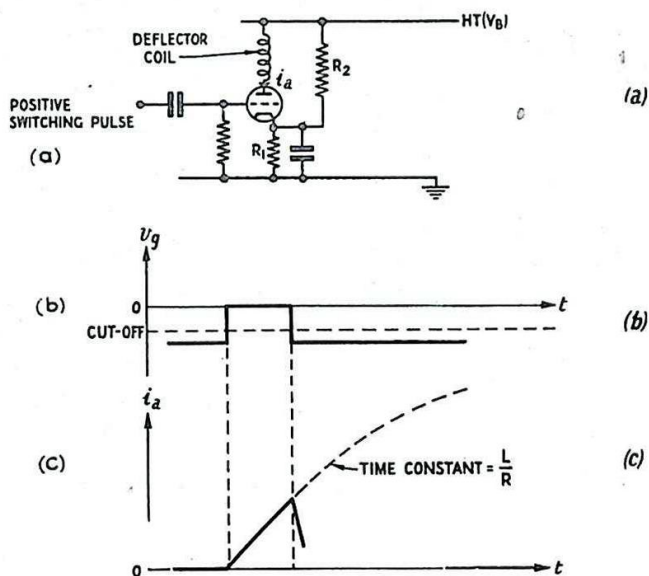


Fig. 558 - Simple time-base generator for magnetic deflection. Application of rectangular voltage pulse.

$$i = \frac{V}{R} \left(1 - e^{-\frac{Rt}{L}}\right),$$

where V is the magnitude of the applied voltage (see Chap. 2, Sec. 5). Provided only a small fraction of the total change of current is utilised the resulting time base is approximately linear (compare Sec. 6).

Fig. 558 shows the circuit of a simple time-base generator. Here the time constant $\frac{L}{R}$ of the circuit (where R is the total

resistance i.e. the sum of the valve resistance and the resistance of the coil) is likely to be long compared with the time during which the valve is allowed to conduct. The valve current is normally cut off by the biasing arrangement which in this case consists of resistors R_1 and R_2 . A positive-going rectangular pulse (b), applied to the grid, renders the valve conducting for the duration of the time base, and the current through the coil increases exponentially.

Consider the effect of applying a linear rise of voltage to a circuit containing L and R in series. If the rate of rise of applied voltage is m , then the current at any instant is given by :-

$$mt = \frac{L di}{dt} + iR \quad (\text{see Chap. 2, Sec. 13}).$$

The solution of this equation is :-

$$i = \frac{mt}{R} + \frac{mL}{R^2} \left(1 - e^{-\frac{Rt}{L}}\right),$$

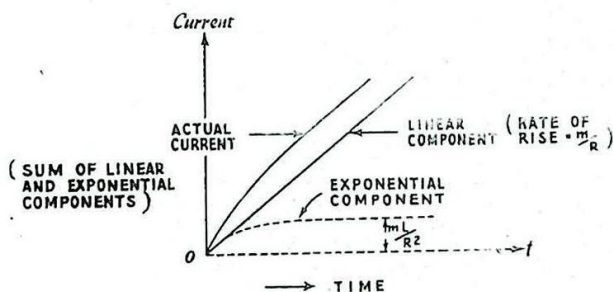


Fig. 559 - Effect of applying a linear rise of voltage to a series L - R circuit.

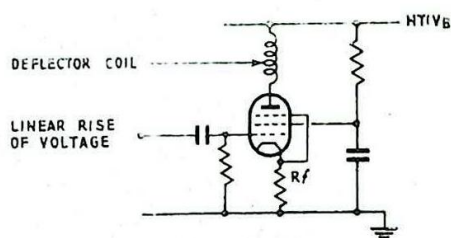


Fig. 560 - Simple time-base generator for magnetic deflection. Application of a linear rise of voltage.

i.e. the current consists of two components one of which varies linearly with time and the other exponentially (Fig. 559). If the time constant of the exponential rise is small in comparison with the duration of the applied voltage, the exponential component quickly rises to its maximum value; subsequently the rise of current is linear. For a given value of L the time constant $\frac{L}{R}$ is made smaller by increasing the value of R . However, such an increase in R reduces the linear rate of rise of current, and therefore decreases the total rise of current taking place in a given time interval.

A suitable circuit arrangement is shown in Fig. 560. The deflector coil is shown in the anode circuit of a pentode. A linear rise of voltage, developed by one of the circuit arrangements described

in Sec. 7, is applied to the grid of the pentode, which is biased so that it operates over the linear portion of its characteristic. The total resistance in series with the inductance is the sum of the resistance of the coil and the effective slope resistance of the valve. The effective slope resistance of the valve is increased, and therefore the time-constant $\frac{L}{R}$ is decreased, by the provision of negative current feedback, which is due to the presence of the un-bypassed resistor R_f in the cathode circuit.

In the two arrangements considered so far it is clear that the change of current through the deflector coil is not linear, and can be made approximately so only by the sacrifice of a considerable portion of the possible current change. A more economical method of obtaining a linear change of current is illustrated in Fig. 561.

The current variation is shown at (a). The manner in which the current decreases during interval (1) is usually unimportant, since this interval corresponds to the flyback. It is usually important, however, that the current should increase from zero or some other specified value at the beginning of interval (2). During the interval (2) the current is given by

$$i = at; \quad (b)$$

The corresponding voltage across the resistor is

$$Ri = atR.$$

The voltage across the coil is

$$L \frac{di}{dt} = aL; \quad (c)$$

Hence the total voltage across coil and resistor is $aL + atR$. It follows that the applied voltage during the growth of current is required to be of the form

$$V = mt,$$

where

$$\frac{V}{m} = \frac{L}{R}.$$

Since the manner in which the current returns to zero is unimportant, the shape of the negative-going portions of the applied voltage is usually immaterial.

A voltage variation of the form shown in Fig. 561 (d) is commonly produced by circuits of the relaxation oscillator or relay type. Alternatively, the circuit of Fig. 562(a) can be used. This circuit is essentially a time-base generator of the type described in Sec. 6, but is modified by the inclusion of a resistor R in series with the time-base condenser C . A negative pulse is applied to the grid of the valve so that the valve current is cut off. The rise of voltage across the resistor R is given approximately by:-

$$V = \frac{R}{R_g + R} \cdot V_B.$$

The voltage across the condenser C cannot change instantaneously so that the output voltage of the circuit at instant (1) rises by the same amount.

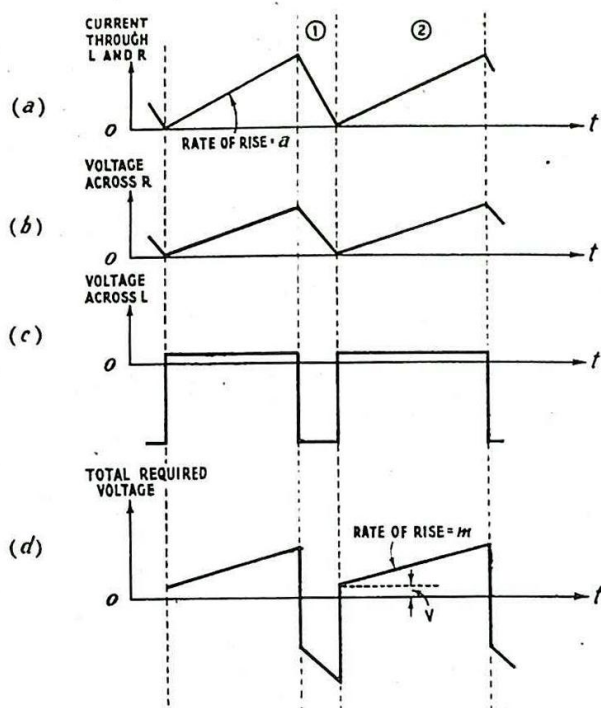


Fig. 561 - Synthesis of voltage required to be applied to a series L-R circuit to obtain a linear rise of current.

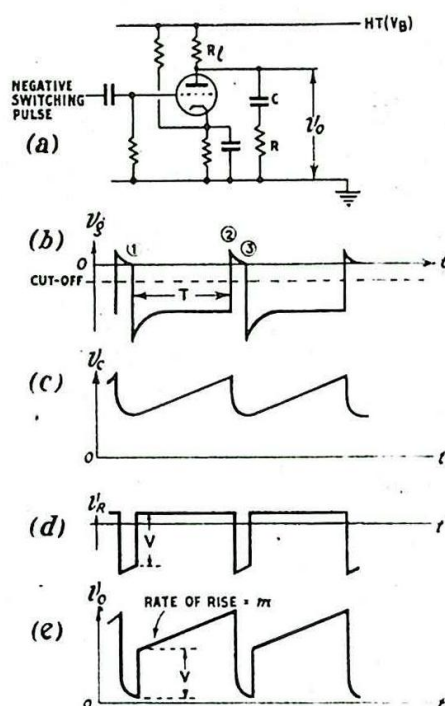


Fig. 562 - Action of time-base circuit which provides an output voltage suitable for producing a linear rise in current if applied to a series L-R circuit.

During the time interval (1) to (2) the condenser charges exponentially with a time constant $C(R_e + R)$. Provided $C(R_e + R) \gg T$, the rise of voltage across the condenser is nearly linear, and the current flowing through R is almost constant. Consequently the output voltage varies as shown at (e).

At instant (2) the valve conducts. Provided R_e is large compared with the resistance of the valve when conducting v_o falls almost to zero. During the interval (2) to (3) C discharges through R and the valve in series.

If the amplitude of the instantaneous rise of output voltage at instant (1) is V and the rate of the linear rise is m the values of V and m depend on the values of R_e and R . It can be shown that variation of R alters the ratio V/m .

It follows from the discussion of the waveforms of Fig. 561 that, if the output voltage of the circuit of Fig. 562(a) is applied to a series L-R circuit to generate a linear time-base current, R must be adjusted so as to make

$$\frac{V}{m} = \frac{L}{R}.$$

A suitable circuit arrangement for producing a

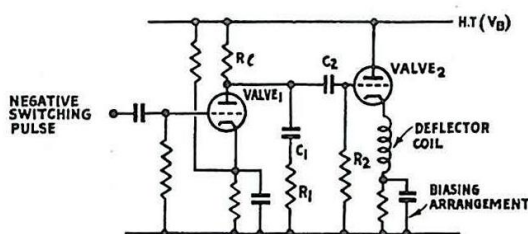


Fig. 563 - Circuit for producing linear increase of current in deflector coil.

linear increase of current in a deflector coil is shown in Fig. 563; this arrangement incorporates the circuit of Fig. 562. Valve 2 can be biased so that little if any current flows through the coil, except during the appearance of the time base, so that the deflection of the spot starts from the centre of the screen. Such an arrangement is advantageous if a radial time base (Sec. 13) is required. In this circuit the effective resistance of the L-R circuit includes the anode-cathode resistance of the valve 2 and the resistance of the coil.

Provided the time-constant C_2R_2 is large compared with $C_1(R_e + R_1)$, little distortion is introduced by the coupling circuit so that the output of the cathode follower applied to the L-R network is an accurate reproduction of the voltage across C_1 and R_1 .

When a valve circuit is used for the generation of time base currents for magnetic deflection it is usually necessary to produce the fly-back by cutting off the current of the time-base valve, which has the deflector coil in its anode or cathode circuit. This causes the coil to ring with its self-capacitance and strays, and unless precautions are taken this ringing may have deleterious effects. The anode voltage may be raised to such a high value as to damage the valve, or to render it conducting again. Further, the oscillations may be prolonged so that they interfere with the succeeding time-base sweep (Fig. 564). The oscillations may be damped out by the inclusion of a series or parallel damping resistor of suitable value.

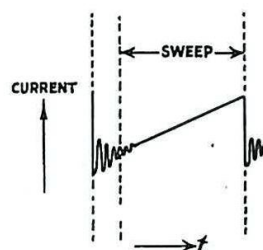


Fig. 564 - "Ringing" of deflector coil.

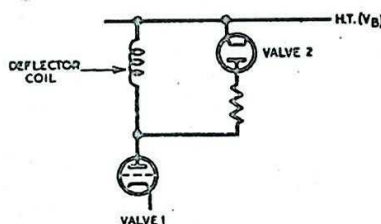


Fig. 565 - Damping circuit.

Such a resistor may affect the manner in which the current through the coil changes during the time-base sweep. It may be replaced by a diode and resistor together in parallel with the deflector coil (Fig. 565). When the current through the valve is increasing, i.e. when the time-base is being formed, the voltage across the coil holds the anode voltage of valve 1 below HT. Consequently, the diode does not conduct since the voltage at its anode is negative with respect to that at the cathode. When the valve current is suddenly cut off the anode voltage of this valve, and therefore the anode voltage of the diode, rises above that of the HT supply. The diode conducts and the energy of the electromagnetic field of the coil is dissipated in the resistor.

12. CIRCULAR, ELLIPTICAL AND SPIRAL TIME BASES

One form of time-base generator is an oscillator or other source of sinusoidal oscillations. If a sinusoidal voltage is applied between a pair of deflector plates of a CRT a straight line time base is produced. A similar result is obtained if a sinusoidal current passes through a pair of CRT deflector coils. Under such circumstances, the CRT spot moves to and fro with a simple harmonic motion, and the displacement of the spot from its rest position at any instant is proportional to $\hat{v} \sin \omega t$ (where \hat{v} is the peak value of the deflecting voltage and angular frequency $\omega = 2\pi f$). If a sinusoidal voltage (or

current) of large amplitude is used, the effective time-base voltage is practically linear (Fig. 566); however, the spot remains on the screen for only a small fraction of the period of the sinusoidal voltage and the time base is useful only if the waveform to be examined is of comparatively short duration. The fly-back may need blacking out since it traverses the screen at the same rate as the forward sweep.

A time base in the form of an ellipse or circle is formed on the screen of a CRT if one sinusoidal voltage is applied between one pair of deflector plates, and another sinusoidal voltage of the same frequency in quadrature with the former is applied to the other pair. This assumes that the deflection sensitivity for each pair of plates is uniform in the region of the trace. If the deflections produced by the two voltages are equal in amplitude the time base is circular in form, otherwise it is elliptical. As the frequencies of the two sinusoidal voltages must be the same these voltages are usually derived from the same source, the phase difference being introduced by means of a condenser-resistor arrangement; (Fig. 567). In the case of the circular time base the spot moves at a constant speed (i.e. the time base is linear) and the trace is useful for the whole of its period.

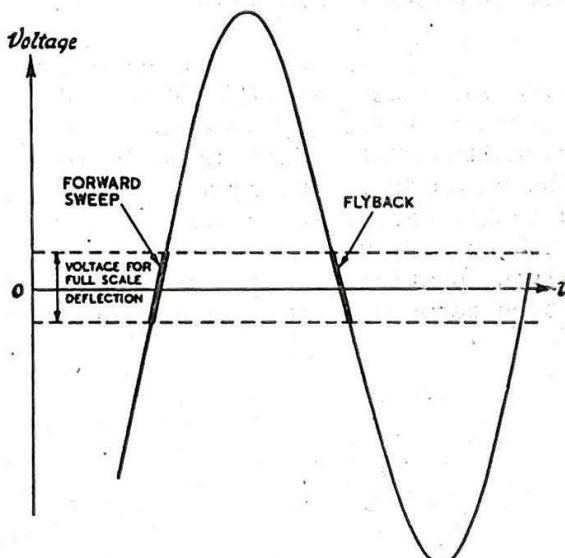


Fig. 566 - Sinusoidal time-base voltage.

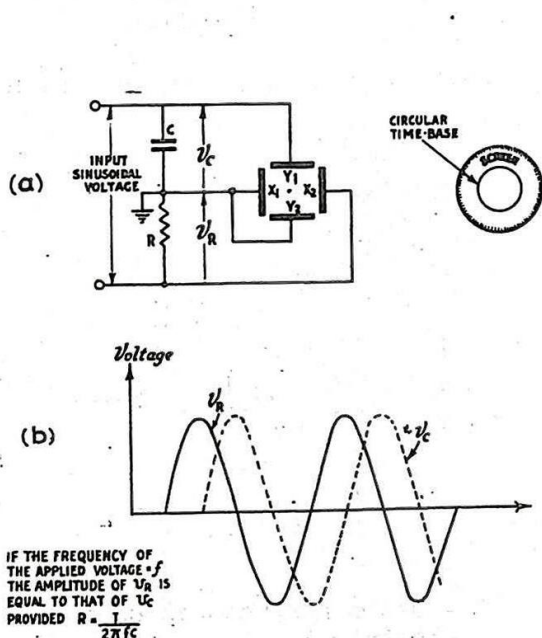


Fig. 567 - Production of a circular time-base. Electric deflection.

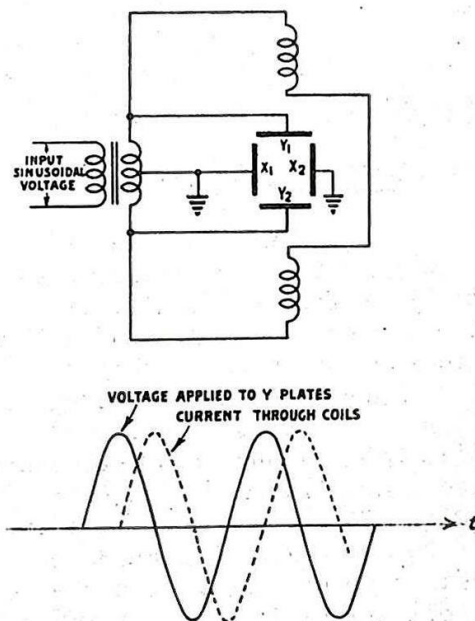


Fig. 568 - Production of a circular time-base. Mixed deflection.

An elliptical or circular time base may also be produced using electric and magnetic deflection. Consider the arrangement of Fig. 568; the vertical magnetic field gives the horizontal deflection while the electric field of the Y-plates gives the vertical deflection. The current through the deflector coils is in quadrature with the supply voltage, and therefore with the voltage applied to the Y-plates, so that no extra phasing circuit is required. The X-plates are available for other purposes.

In the circuits which have been described for producing circular time bases the effect of stray capacitance has been ignored; in practice the input capacitance of deflector plates and the self capacitance of deflector coils are appreciable and modifications to the circuits are necessary. For a CRT of given screen diameter a circular time base provides a trace of greater length than is possible with one of straight line form. Therefore, if the time base is of a given duration, the circular shape gives the greater visual discrimination between signals separated by small intervals of time.



Fig. 569 - Spiral time-base.

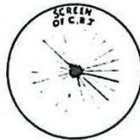


Fig. 570 - Radial time-base.



Fig. 571 - PPI display.

If the time-base voltages applied to the CRT are not truly sinusoidal, but decrease in amplitude from one cycle to the next, the time base takes a spiral shape; (Fig. 569). Such a shape allows for a still greater extension of the length of the time base. A damped ringing circuit is suitable for generating this kind of time base. Alternatively, a sinusoidal voltage may be applied to an amplifier, to which is fed a sawtooth modulating voltage, thereby producing a time-base voltage of the required form. Suppose that the ratio of the frequency of the sinusoidal voltage to that of the sawtooth generator is r . If r is an integer, a stationary time base is formed. If r is not exactly an integer the time base will appear to rotate.

If $\frac{1}{r}$ is a large integer a radial time base is formed which looks like the spokes of a wheel (Fig. 570).

13. RADIAL TIME BASE

If a time base of straight line form is arranged to start from the centre of the screen of a CRT, and is then made to rotate about this point at a rate which is slow compared with the repetition rate of the time base, a number of successive radii are traced on the screen; (Fig. 570). The resulting trace is termed a Radial Time Base.

This type of time base has wide applications in radar, forming the basis of the Plan Position Indicator (PPI) display. The electron beam of the CRT is

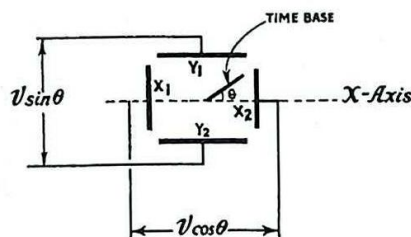


Fig. 572 - Arrangement used in electrical method of producing a radial time base.

normally modulated by signal voltages, so that the presence of signals is indicated by a brightening of a portion of the trace (Chap. 6, Sec. 31). The brilliance of the trace as a whole is adjusted so that only those portions are visible which have been brightened by signal voltages. Provided the screen of the CRT is of the after-glow type, bright patches indicating the presence of target echoes can be visible at different parts of the screen simultaneously; (Fig. 571). One method of producing such a time base, which however is not normally used in radar, has already been indicated at the end of Sec. 12.

The number of spokes formed on the screen of the CRT is given by the ratio:-

$$\frac{\text{Period of Rotation}}{\text{Repetition Period of time base}}$$

and if this is of the order of a 1000, the screen is likely to have almost every part of its surface covered by the trace during each period of rotation.

There are two main methods used to produce a radial time base. The first method, which is sometimes used if the time base is developed by magnetic deflection, is to rotate mechanically the deflector coils about the axis of the CRT. Such an arrangement works well for low speeds of rotation but at high speeds mechanical problems become difficult, and it is preferable to employ a system which avoids the use of moving parts at the CRT. In any case, if the time base is developed by electric deflection, mechanical rotation of the deflector plates is not easily obtained.

The basis of the electrical method of producing a radial time-base is as follows. Two similar sawtooth voltages are produced, one of which has an amplitude $V \sin \theta$ and the other an amplitude $V \cos \theta$. If these two voltages are applied to the Y- and X-deflection systems of a CRT respectively, the straight-line time base produced makes an angle θ with the X-axis, (Fig. 572). As θ is made to vary, the time base rotates.

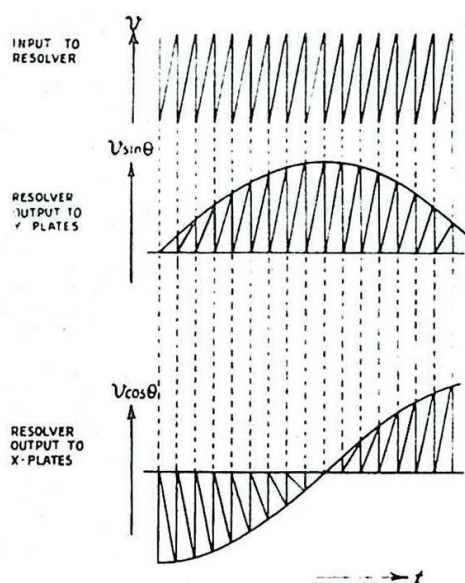


Fig. 573 - Resolver voltages.

The voltages necessary for producing a radial time base are normally derived from a voltage of recurrent sawtooth waveform and of constant amplitude. Such a voltage is applied to a resolver which may take any one of the forms described in Chap. 3. The resolver provides two output voltages each similar to the input but with amplitudes proportional to $\sin \theta$ and $\cos \theta$ respectively where θ is the mechanical setting of the resolver (Fig. 573).

CHAPTER 12

RECTIFYING AND CLAMPING CIRCUITS

1. RECTIFICATION

A description of the process of rectification (and detection) is given in the Radio books of three Services. However, certain principles of rectification are discussed here, since they form the basis of a knowledge of Clamping, and it is desirable that the connection between the two processes should be clear.

A rectifying or detecting circuit is shown in Fig. 574(a). Assume for simplicity that the applied voltage v_i has the rectangular waveform shown in Fig. 574(b) and that its amplitude is denoted by \hat{v}_i .

At the start of the operation the condenser C is uncharged and the voltage across it is therefore zero. At the beginning of interval (1) the applied voltage rises suddenly from zero to \hat{v}_i . This change of voltage appears instantaneously across the diode, which then conducts heavily. Hence C charges through the diode, and the voltage across it rises with a time-constant CR_D towards \hat{v}_i (Fig. 574(c)). (R_D is the resistance of the diode when this valve is conducting.) During interval (1) the voltage across the condenser rises to V' . Simultaneously, the voltage across the diode falls to $(\hat{v}_i - V')$ (Fig. 574(d)), since at any instant the sum of the voltages across the condenser and across the diode must equal the applied voltage.

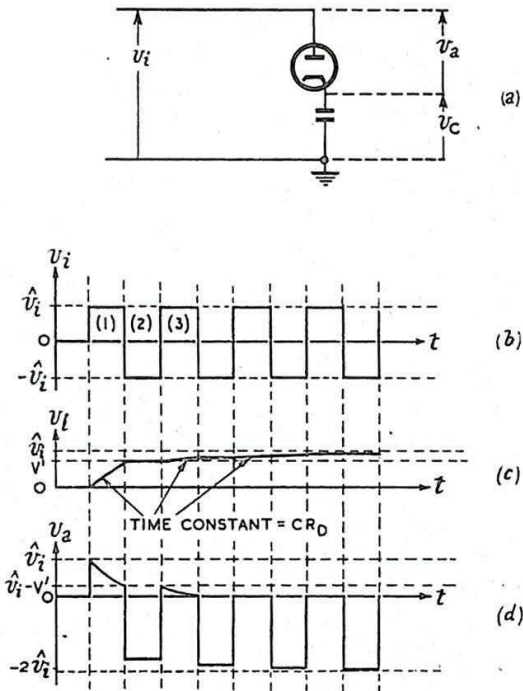


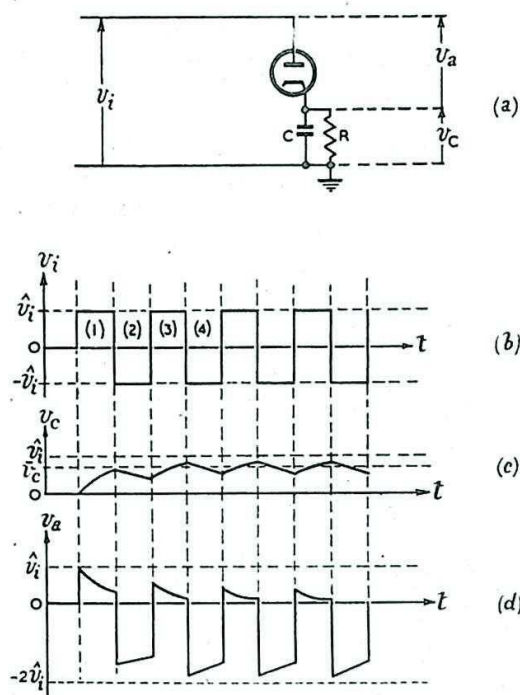
Fig. 574.- Action of simple rectifying circuit.

At the end of interval (1) the applied voltage falls suddenly by an amount $2\hat{v}_i$. This voltage change appears instantaneously across the diode, and consequently the current is cut off. Since no current flows in or out of the condenser C , the voltage across C remains constant. So long as the applied voltage remains steady (interval (2)), the voltage across the diode does not vary.

At the beginning of interval (3) the applied voltage, and therefore the voltage across the diode, rises suddenly by an amount $2\hat{v}_i$. The diode conducts as soon as the input voltage across it reaches V' . Consequently, the voltage across the condenser rises exponentially, with time-constant CR_D from V' towards \hat{v}_i (interval (3)). The rise of voltage across the condenser C during interval (3) is not so great as during the equal interval (1), since the voltage range over which the condenser tends to charge is smaller. The voltage across the diode falls during interval (3) by the same amount as the voltage across the condenser rises.

The voltage v_c across the condenser eventually reaches a value \hat{v}_i at which it remains. The voltage across the diode then has a mean value of $-\hat{v}_i$ i.e., the peak voltage is zero. A similar result is obtained if the applied voltage is sinusoidal.

In practice the voltage developed across C is the output of the rectifier circuit, which is normally used as a source of steady voltage for other circuits. The load on the rectifier due to these other circuits taking current can be represented by a resistor R connected in parallel with C, (Fig. 575(a)). Under such circumstances the operation of the circuit is modified since C can discharge through R.



(b) Fig. 575.- Action of practical rectifying circuit.

Assuming, as is normal with a rectifying circuit, that R is large compared with R_D , C charges with time-constant approximately equal to CR_D , during the time the diode conducts. When the diode is not conducting C discharges with time constant CR. As in the case of an alternating voltage applied to a series C-R circuit (see Chapter 2 Sec.4) the circuit eventually settles down to a steady state in which the charge accumulated on C during each cycle is equal to the charge which leaks away through R (Fig. 575(c)). From then onwards the voltage across the condenser fluctuates, during successive cycles, about a constant value \bar{v}_c , slightly smaller than \hat{v}_i . The greater is the value of R compared with R_D the more closely does the mean value of the voltage across the condenser approach to \hat{v}_i , and the smaller the amplitude of the fluctuations about this mean value.

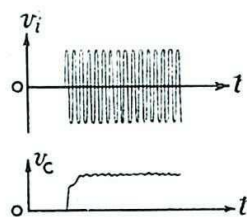


Fig. 576.- Detection of CW oscillations.

The above discussion applies also to the detection of amplitude-modulated CW. Fig. 576 illustrates a CW oscillation which is to be detected by the circuit of Fig. 575a. Provided the time-constant CR is much longer than the period of oscillation the steady value of the voltage across the condenser is almost equal to the amplitude of the applied voltage. If the oscillation (frequency f_0) is amplitude

modulated (modulation frequency f_m), as shown in Fig. 577, then, provided $\frac{1}{f_o} \ll CR \ll \frac{1}{f_m}$

the voltage across the condenser, at any instant, conforms almost exactly to the shape of the modulation envelope.

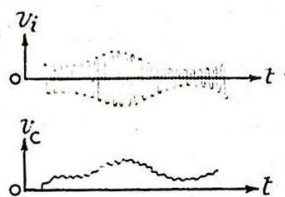


Fig. 577.- Detection of amplitude modulated CW oscillations.

In radar equipments the voltage applied to the detector circuit takes the form of radio-frequency pulses as shown in Fig. 578. Such a voltage may under certain circumstances be considered as an extreme form of that shown in Fig. 577. If the voltage across the condenser is to conform faithfully to the pulse shape it is necessary that $\frac{1}{f_o} \ll CR \ll T$, where T is the duration of the pulse.

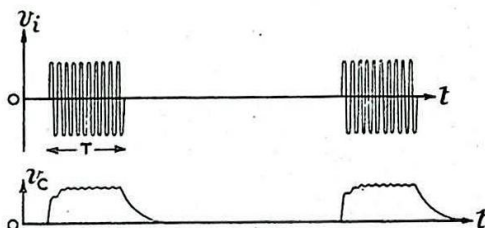


Fig. 578.- Detection of RF pulse.

2. SIMPLE CLAMPING CIRCUITS

So far, in the discussion of the operation of the circuits shown in Figs. 574 and 575, the emphasis has been on the form of the voltage across the condenser. However, it has been pointed out that the voltage developed across the diode is almost entirely in one sense only. If one of the electrodes is connected to a source of constant voltage the voltage at the other electrode is prevented from either rising above or else falling below this fixed level according to the method of connection. This process is known as Clamping. Various other terms are used to describe clamping, depending on the nature of the input voltage variations.

If the input consists of a succession of unidirectional pulses, either all positive-going or all negative-going, and the output is clamped so that the base-line is at a fixed level independent of the amplitude of the pulses, the process is called Base-line Stabilisation. If, on the other hand the peaks (of either the positive-going or the negative-going positions) are clamped, the term Peak Level Stabilisation is used. The term DC Restoration also is frequently used to describe these processes, but strictly should be reserved for the case in which the mean level of the output voltage is made the same as that of the input. Common examples of radar circuits utilising peak-level stabilisation are those employing valve switching or "gating" pulses or slide-back biasing arrangements (Chapter 7). Where gating pulses are applied to release the bias on either the suppressor or the control grid of a valve it is usually necessary to ensure that the level at which the grid is held during the conducting period should be clearly defined. This may be achieved by grid current flow, but, the addition of a diode usually facilitates clamping. Examples of this form of clamping are to be found in Chapters 10 and 11.

Base-line stabilisation has common application to CRT circuits. For example, in PFI systems, where the signal pulses are usually applied to the grid or cathode via a C-R network, it is important that an increase in the amplitude of some signals should not depress the general level of

the brightness and so make weaker signals ineffective. If the base-line is clamped at a predetermined level, the amplitude of signal necessary to cause an indication on the screen is constant. Clamping is also commonly employed in time-base circuits and in deflection systems using a simple A-display. Consider the latter case, in which the signal pulses are applied via a C-R circuit to the Y-plates to produce the trace shown in Fig. 579. An increase in the mean signal amplitude would, in the absence of any clamping device, cause the base-line to be depressed. If this is undesirable, a clamping diode is usually employed to ensure that the base-line occurs at a fixed position on the screen.

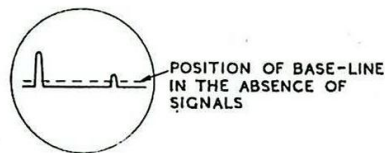


Fig.579.- Shift of base-line in presence of signal.

The simplest type of clamping circuit is shown in Fig.580. So far as the voltages across C and the diode are concerned, the operation of this circuit is essentially the same as that of the circuit shown in Fig.575a

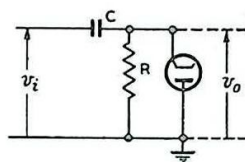


Fig.580.- Simple circuit for clamping negative-going excursions of output voltage.

Consider first the effect of applying to this circuit in the absence of the diode the input voltage variation shown in Fig.581a (This is the type of output commonly obtained from the anode of a pulse-amplifying valve). Provided the time-constant CR is long compared with the repetition period of the input pulses, the output voltage in the steady state takes the form shown in Fig.581(b). The voltage developed across C is shown at (c); (compare Chapter 2 Section 4).

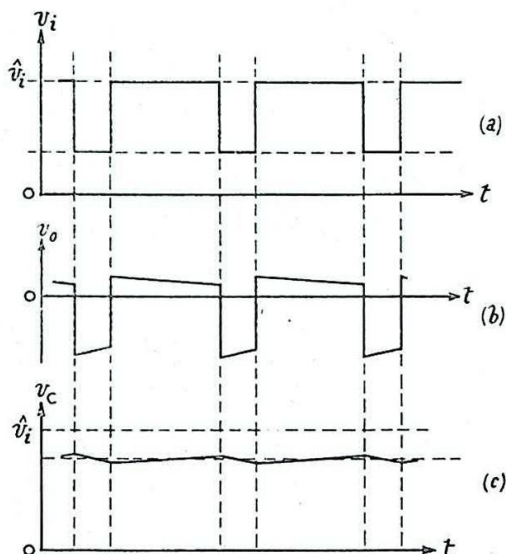


Fig.581.- Action of circuit of Fig.580; diode out of circuit.

If the voltage variation of Fig.581(a) or 582(a) is applied to the circuit of Fig.580, this time with the diode in circuit, the circuit conditions settle down to the state depicted in Fig.582(b) and (c). In the steady state the peaks of the negative-going pulses are clamped at zero potential. The condenser is charged to approximately \hat{v}_i . Since $CR \gg T$, only very little additional charge is acquired while the input is at \hat{v}_i , the diode being non-conducting, so that the condenser

voltage v_c rises by only a small amount. When the input falls again to \hat{v}_i the diode conducts and the additional charge rapidly leaks away and v_c falls again to \hat{v}_i .

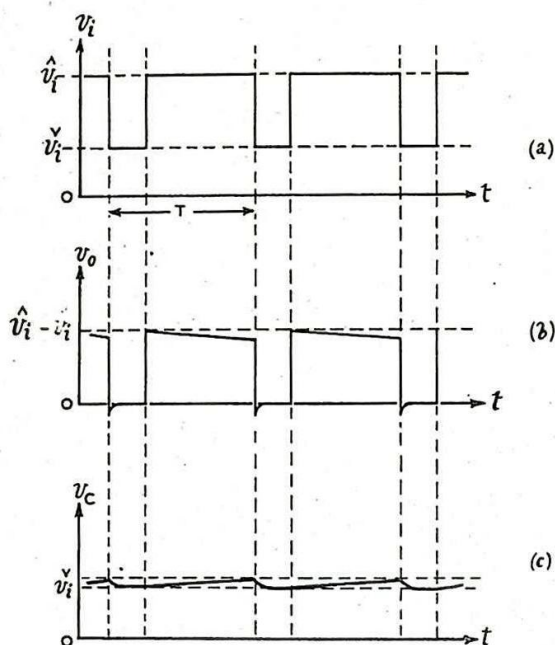


Fig.582. - Action of circuit of fig.580; diode in circuit.

The effect of such a circuit in practice may be even more pronounced than has already been indicated, owing to the low value of the input resistance of the circuit when the diode is conducting. This occurs in the arrangement shown in Fig. 583(a). In the absence of the diode, the output voltage shown at (c) is developed across R in response to the input shown at (b). When the diode is in circuit, the output voltage settles down to the steady state shown at (d).

The action of the circuit of Fig. 583(a), with the diode inserted, is as follows.

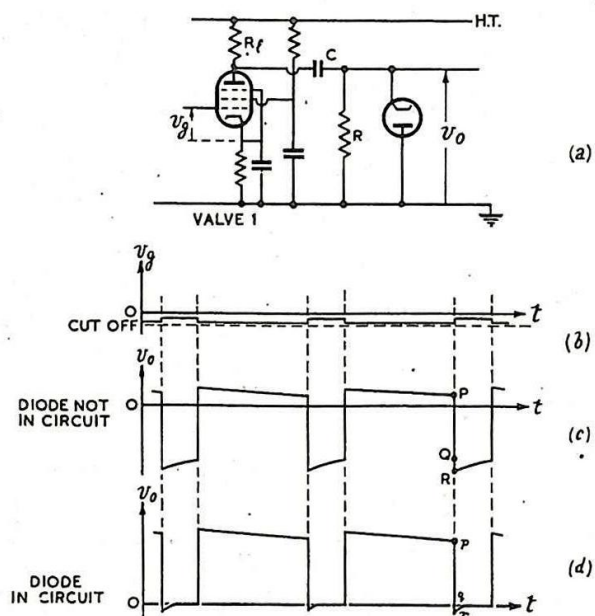


Fig.583. - Effect of low input resistance of diode.

While the diode is not conducting no change in the amplitude of the output voltage variation occurs, so that p_q (d) is equal to PQ (c). During the conducting period the input resistance to the clamping circuit is very small (R_D), and since this is effectively in parallel with the anode load the amplification is substantially reduced. This causes the negative-going peak q_r (d) to be very much less than the corresponding peak QR (c), so that the clamping effect is considerably enhanced.

Fig. 584a shows the diode circuit in a form suitable for clamping the positive-going extremities of the output voltage. The steady state conditions are illustrated in Fig. 584. The condenser charges during the positive-going excursions of the input to approximately \hat{v}_i , discharging slightly during the negative-going excursions. This causes the output voltage to rise slightly above zero potential at the beginning of each positive-going excursion, so that the diode conducts, and rapidly recharges the condenser to \hat{v}_i .

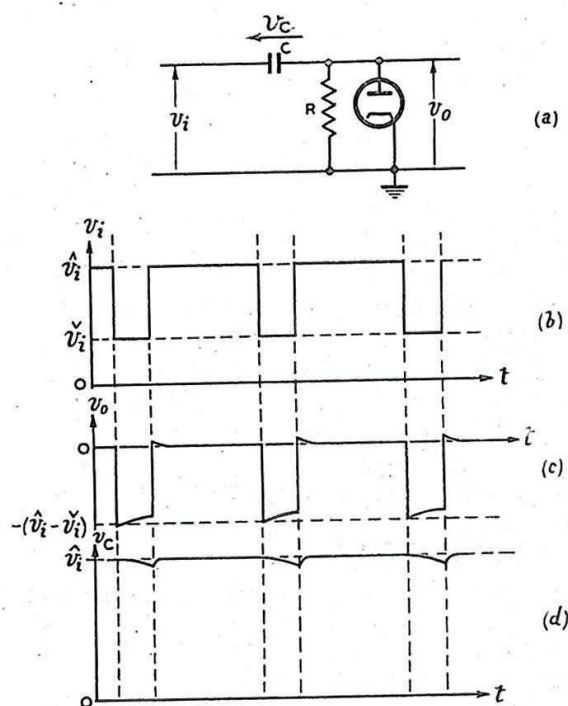


Fig. 584.- Clamping of positive-going excursions of output voltage.

Clamping may be performed at the grid of a triode or pentode. Thus the grid of the valve shown in Fig. 585 acts as the anode of a diode; any tendency for the grid voltage to rise above zero causes grid current to flow and the condenser C is charged rapidly. The arrangement shown is essentially that needed to obtain automatic bias by using the flow of grid current (see Chapter 7 Section 4). Clamping is an extreme form of such automatic biasing.

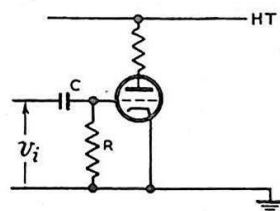


Fig. 585.- Use of control grid for clamping

The circuits discussed above clamp one or other extremity of the output voltage to zero. However, clamping may occur at any other desired reference level. For example, Fig. 586(a) shows a circuit which is capable of clamping the upper extremity of the output at some positive level V . The steady state conditions of the circuit are illustrated at (b) and (c)

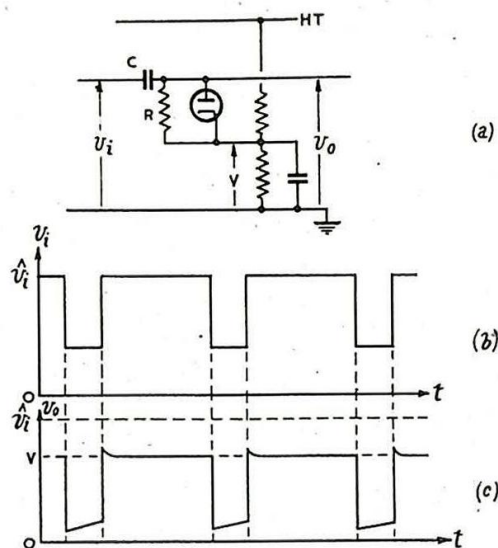


Fig.586.- Clamping at a level other than zero.

Clamping may also be applied to a time-base voltage. The necessity for such a procedure depends on the type of voltage variation and on the conditions under which it is used.

Fig. 587(a) shows a typical time-base voltage which has a discontinuous sawtooth waveform. If this voltage is applied, say, to a deflector plate of a CRT, via a C-R coupling circuit, the deflector plate voltage varies about a mean level represented by the line AB at (b). This mean level is usually determined by the setting of the shift control. It is often necessary, in radar apparatus, to change the duration of the time base, without altering the repetition frequency (Fig.587(c)). Figs.587(b) and (c) show how the change in time-base duration alters the level from which the time-base voltages rise, and so changes the starting position of the time-base on the screen of the CRT. Similarly, if there is a change in the amplitude of the time-base voltage as shown at (d), there is again a change in the starting position of the time-base. Where changes of time base duration or amplitude are necessary it is clearly desirable to introduce clamping.

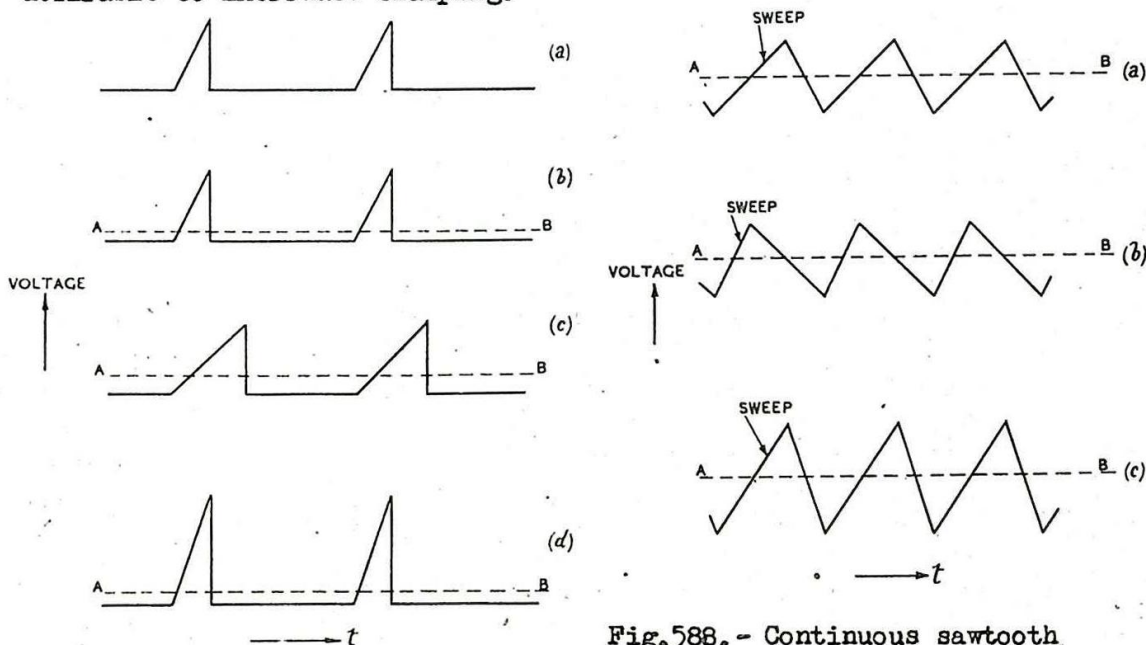


Fig.587.- Sawtooth pulses; effect of changing amplitude or duration.

Fig.588.- Continuous sawtooth voltage; effect of change of amplitude or sweep-to-flyback ratio.

If the time-base voltage has a continuous sawtooth waveform, as shown in Fig. 588(a) conditions are rather different. For example, (a) and (b) represent time-base voltages of the same amplitude but with different sweep-to-flyback ratios. In these cases there is no difference between the voltage levels from which the time-base voltages rise. However, if there is a change in the amplitude of the time-base voltage, as illustrated at (c), there is a change in the starting voltage level, and therefore in the position of the start of the time-base on the screen of the CRT.

This may not be deleterious, since the mean value of the time-base voltage corresponds to the centre of the trace, which is usually positioned at the middle of the screen. The increase in amplitude causes the time-base to extend farther across the screen in both directions.

It may be desirable for the time-base to start from the centre of the screen of the CRT (e.g. PPI presentation), i.e., for the effective time-base voltage to rise from zero. This may be accomplished by clamping the lower level of the time-base voltage to zero, by one of the methods described in this chapter.

3. SWITCHED CLAMPING CIRCUITS

The circuits described in Section 2 are examples of a simple type of clamping in which one extremity of the output voltage is clamped at a fixed level. However, sometimes a different type of clamping is required, in which both positive-going and negative-going output pulses are made to start from some fixed level.

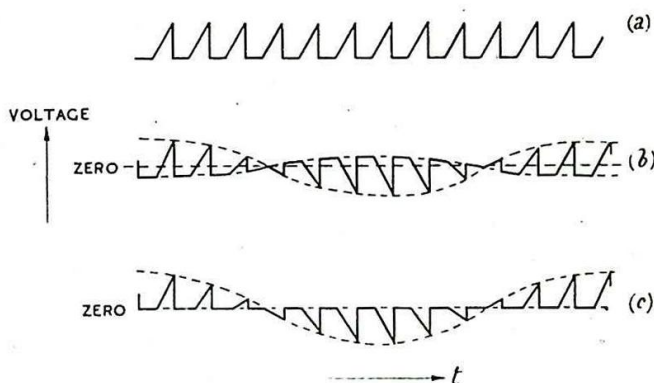


Fig. 589.- Voltage variation requiring switched clamping.

In Chapter 11 Section 13 a method of producing a rotating time-base is described. An essential part of the method is the use of a resolver. A sawtooth voltage as in Fig. 589(a), is applied to the resolver and if a non-resistive resolver is used, output voltages of the form shown at (b) are obtained. The steady component of the applied voltage is lost in the resolver and each output pulse settles down about its mean value. As the sawtooth pulses change in amplitude the time-base voltages start from varying levels, and not from zero as is required. If all the time-bases are to start from the centre of the screen the output voltage should take the form shown at (c). The transformation of the waveforms from the form of (b) to that of (c) is accomplished by clamping the voltage during each interval between successive sweeps to a fixed value (i.e., zero). This requires that the clamping circuit should be switched into operation during these intervals only.

Fig. 590(a) shows a circuit arrangement for providing switched clamping. The output voltage is maintained at a steady value except for the time during which the clamping valves (valve 1 and valve 2) are rendered non-conducting by means of negative-going pulses, lasting throughout each time-base sweep, which are applied to the control grids of both valves. The relation between the switching pulses and the applied voltage is shown at (b) and (c).

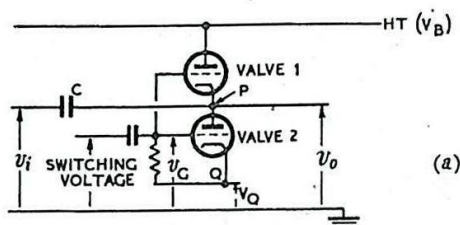
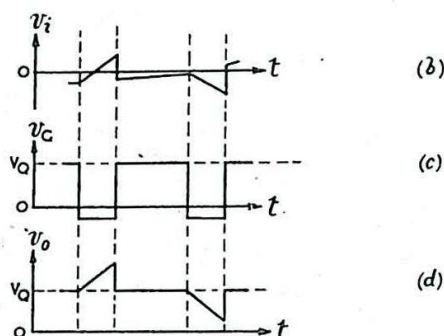


Fig.590.- Action of switched clamping circuit



Valves 1 and 2 are in series between the voltage levels V_B and V_Q (the voltage at Q). While the common grid voltage is held well below V_Q and v_o (which is also the voltage at P), neither valve is conducting the clamping circuit is inoperative. During the positive-going portions of the switching voltage (interval (a)) the grids are held at approximately V_Q , being clamped by the flow of grid current. The voltage at P then settles down to a steady value depending on the relative resistances of the two valves. This steady value cannot be greater than V_Q by an amount sufficient to cut off the current in valve 1. Hence, in the steady state, v_o would be stabilised a few volts above V_Q . As the input voltage v_i rises, v_o tends to rise, reducing the current in valve 1 so that the resistance of this valve is increased. This tends to reduce v_o and so counteracts the rise in input voltage. In the extreme case, when the rise in v_o is sufficient to cut off the current in valve 1, valve 2 acts as a clamping diode so that v_o is prevented from rising substantially above V_Q .

Similarly, if the input voltage falls the resistance of valve 1 decreases and tends to offset the fall in input voltage. If the voltage falls by an amount sufficient to cut off the current in valve 2, valve 1 acts as a clamping diode so that v_o is prevented from falling appreciably below V_Q .

The output voltage can be clamped at any desired level (below V_B) during the positive-going portions of the switching voltage by a suitable choice of the clamping level V_Q . In the application described the switched clamping circuit ensures that all the time-bases start from the same point. A potentiometer arrangement enables V_Q to be determined so that this point is brought to the centre of the screen.

CHAPTER 13

THE PRODUCTION OF SHORT DURATION PULSES

1. INTRODUCTION

A succession of pulses is often required to initiate, at regular intervals, the operation of certain radar circuits (see Chaps. 11 and 14). Normally such pulses have a repetition period of one or two milliseconds and a duration of a few microseconds. This type of voltage variation is not conveniently obtained directly by applying a sinusoidal voltage to a limiting amplifier, for the following reasons. Firstly, there is difficulty in producing, and applying to the limiting amplifier, sinusoidal oscillations of sufficiently large amplitude and sufficiently high frequency to give rise to pulses of short duration with steep sides. Secondly, pulses produced by a limiting amplifier have usually a repetition period which is of the same order as the duration of each pulse: the smaller the mark-to-space ratio, the less truly rectangular are the output pulses (see Chapter 9 Section 6).

It is usually necessary, therefore, to apply the succession of rectangular pulses, either produced by a limiting amplifier or obtained from any other source to some pulse-shaping circuit which converts them into output pulses of sufficiently short duration for the purposes for which they are required.

Various types of pulse-shaping circuits are considered below.

2. THE USE OF A SHORT TIME-CONSTANT C-R CIRCUIT

If a succession of rectangular pulses is applied to a circuit consisting of a condenser C and resistor R in series (Fig. 591), the time-constant being small in comparison with the duration of a pulse, then the voltage developed across the resistor consists of alternate positive and negative pulses, each of short duration, as shown in Fig. 591 (see also Chapter 2).

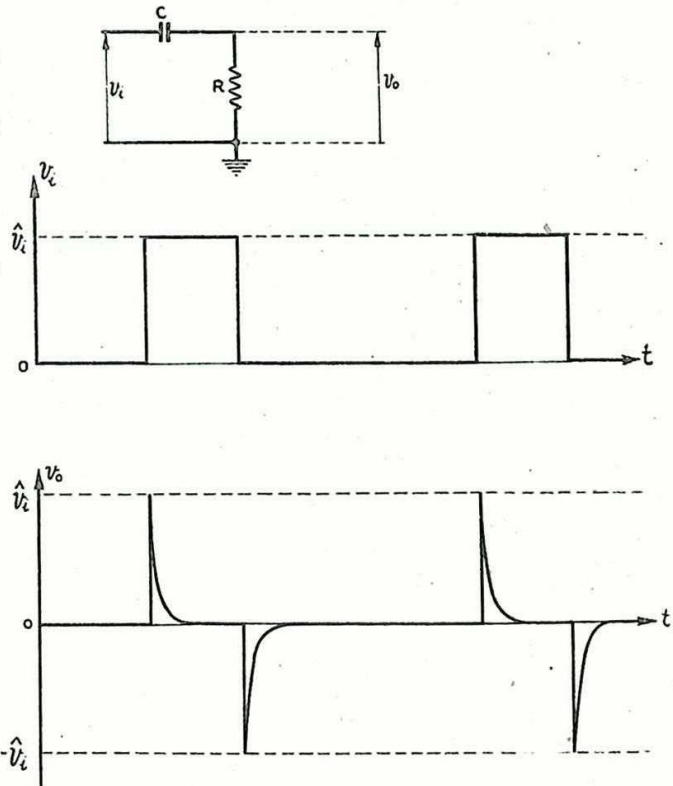


Fig. 591.- Use of short time-constant C-R circuit.

This principle is utilised in the circuit of Fig. 592. Valve 1 is a limiting amplifier which produces a rectangular output voltage from a sinusoidal input. The output voltage of valve 1 is applied to a short time-constant C-R circuit followed by a second limiting amplifier.

This second amplifier may be designed so as to discriminate between the positive-going and negative-going pulses developed across R_g , so that the final output consists of a succession of unidirectional pulses of short duration, (see Chapter 9 Section 8.) In the case depicted this output consists of positive-going pulses. The repetition period of these pulses is the same as the period of the sinusoidal input voltage, as indicated in the figure.

If the voltage variation at the anode of valve 1 could be considered truly rectangular, and if various circuit components other than C and R_g could be neglected, there would be no limit to the reduction in pulse width which could be achieved by diminishing the time-constant CR_g . In practice there are limits beyond which further diminution of the time-constant is not practicable. As shown in Chapter 2 Section 5 and 15, the voltage developed across R_g may be seriously affected by

- (i) the input capacitance C_1 of the succeeding circuit
- (ii) the output resistance of the preceding stage
- (iii) the finite rate of change of the applied voltage - i.e., of the anode voltage of valve 1.

The overall effect of these unavoidable circumstances is to reduce the amplitude of the pulses developed across R_g ; ultimately, if the time-constant is still further reduced, no additional decrease in the duration of the output pulses occurs.

For these reasons, and because they do not produce the rectangular shape of output pulse which is frequently required, C-R circuits usually give place to other types of pulse-shaping circuits when pulses of very short duration, of less than about 5 microseconds, are required. It should be noted that with the circuit arrangement shown in Fig. 593 the rise in the grid voltage of valve 2 at the trailing edge of a negative pulse developed across R_g tends exponentially towards the HT voltage (Fig. 594); it is therefore steeper than when the rise is exponential towards earth potential as would be the case if the circuit of Fig. 592 were used.

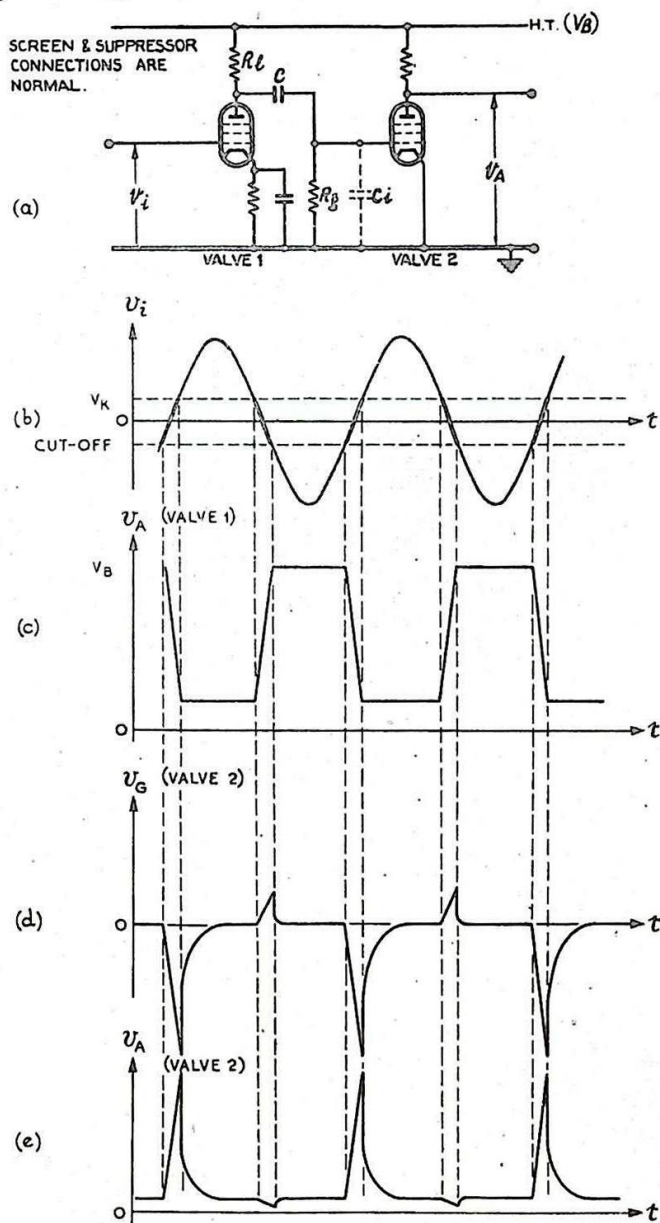


Fig. 592. - Action of limiting amplifier circuit employing short time-constant C-R network.

By this connection of the grid leak to the HT supply instead of to earth the effective pulse width can be substantially reduced without any alteration of the circuit time-constant.

3. THE USE OF A RINGING CIRCUIT

One of the factors limiting the use of a short time-constant C-R circuit for producing pulses of short duration is the input capacitance of the succeeding stage. This capacitance can be utilised in a ringing circuit, as shown in Fig. 595. The response of such a circuit to the output voltage of valve 1 is illustrated in the accompanying diagrams. The frequency of the free oscillation is given approximately by

$$f \doteq \frac{1}{2\pi \sqrt{L(C + C_i)}}$$

Some damping is provided by the output resistance of valve 1 and if this is of suitable value, the parallel damping resistor R may be dispensed with.

The positive-going change of voltage at the anode of valve 1 produces at the grid of valve 2 a damped oscillation commencing with a large positive voltage swing lasting for about a quarter of a period. The negative-going change of voltage produces a similar oscillation commencing with a large negative voltage swing.

The above considerations have neglected any effect of grid current in valve 2. However, in the circuit shown in Fig. 595, since valve 2 is not biased, the positive portions of the damped oscillations are limited. Grid current flows while the grid voltage is positive and the resulting small value of grid-cathode resistance causes considerable damping of the ringing circuit. This increases the logarithmic decrement during the flow of grid current, and reduces in size and duration the subsequent oscillation. As a result, the voltage appearing at the grid of valve 2 consists mainly of

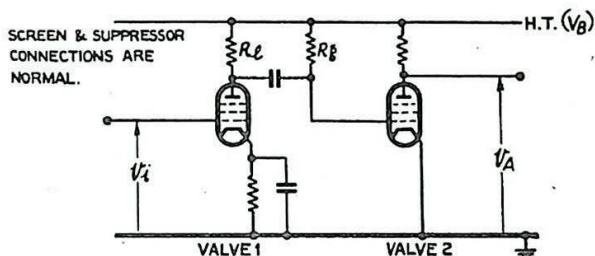


Fig. 593. - Alternative arrangement of the limiting amplifier circuit of fig. 592(a).

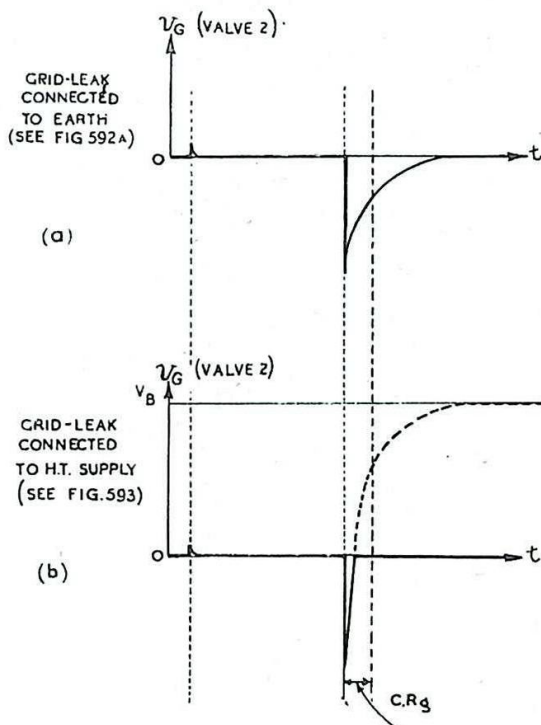


Fig. 594. - Effect of connecting grid-leak to HT supply.

a negative pulse of short duration, the start of which occurs as the anode voltage of valve 1 changes in a negative-going direction.

The duration of pulses produced by this method may be of the order of 1-5 microseconds.

4. THE USE OF A PULSE-FORMING NETWORK

Pulses of short duration, considerably less than 1 microsecond, can be produced from a rectangular input by the use of a pulse-forming network. Fig. 596 shows a possible circuit arrangement. The limiting amplifier produces a rectangular voltage from a sinusoidal input. In the circuit considered here a short-circuited delay network is connected across the anode load. The characteristic impedance of the network is made equal to the load resistance R_L . The operation of the delay line has been discussed in Chapters 2 and 4. In the present case a sudden flow of current in the limiting amplifier produces a rapid change of voltage across R which is in parallel with the input impedance of the network. This voltage is half that which would be obtained if the anode load consisted of R_L alone. The voltage change developed across R_L now travels along the network and after a certain time reaches the short-circuit where it suffers a reflection with a change of sign. This reversed voltage change travels back along the network and on reaching the resistor R_L reduces to zero the voltage developed across it. Since the network is matched to R_L there is no second reflection at this end. A voltage pulse is therefore developed across R_L , as shown in Fig. 597. The duration of

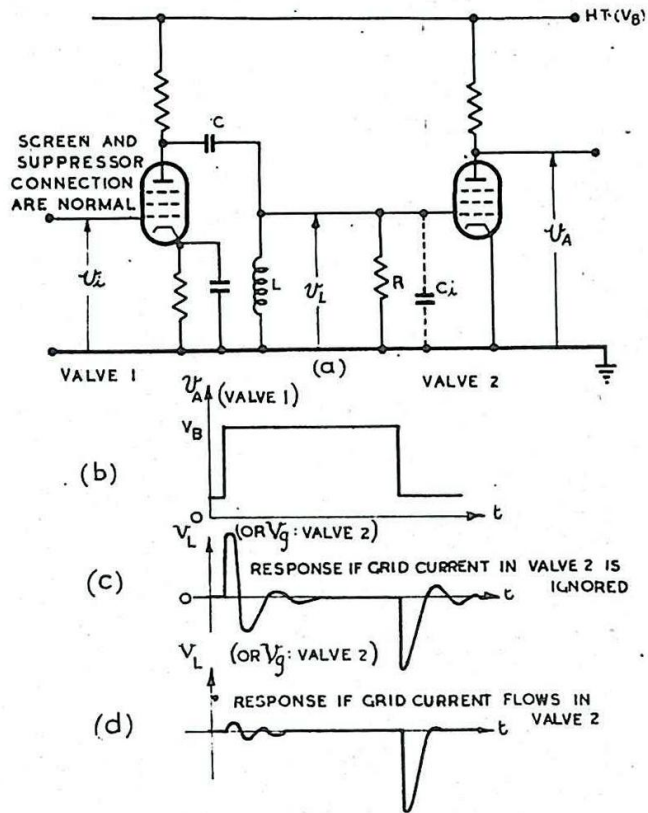


Fig. 595. - Use of "ringing" circuit.

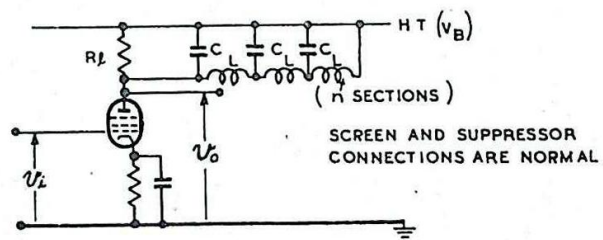


Fig. 596. - Use of delay network in anode circuit.

the pulse is $2n\sqrt{LC}$ where n is the number of sections of the network and L and C are respectively the inductance of each coil and the capacitance of each condenser (see Chapter 3 Section 16). This output pulse is negative-going and its leading edge occurs at the instant at which the current begins to flow through the limiting amplifier valve. If the current through the limiting amplifier is cut off, a positive-going pulse is formed which has a duration equal to that of the negative-going pulse.

A variation of the pulse-forming circuit described above is shown in Fig. 598. Here the network is in the cathode circuit of an amplifier, and is approximately matched at one end by the low output resistance presented by the valve to its cathode load (see Chapter 7 Section 18).

The termination at the other end of the network is a resistor R_t of such a large value that this end can be considered as open-circuited. This resistor is necessary to complete the direct-current path from the cathode of the valve to earth.

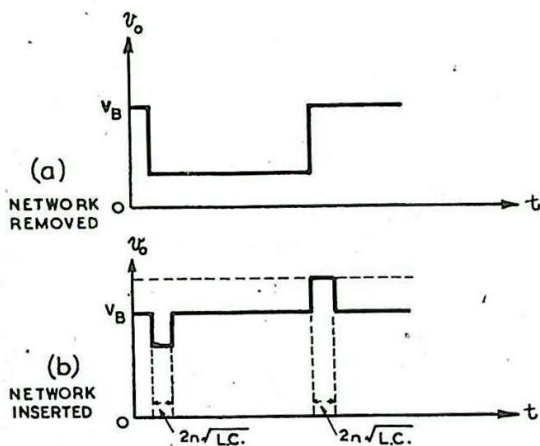


Fig. 597.- Effect of network in anode circuit.

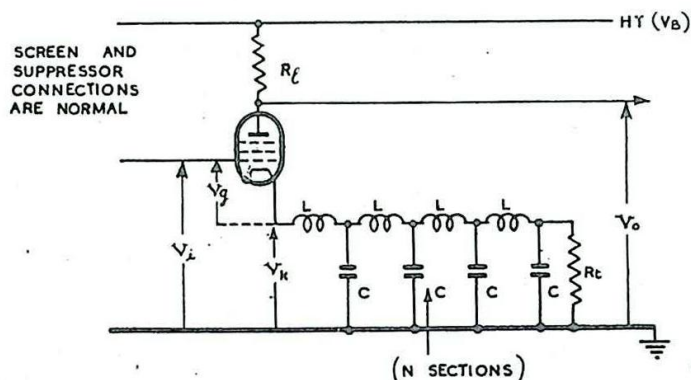


Fig. 598.- Use of network in cathode circuit.

Denote by v_i the input voltage and by v_K the voltage developed between cathode and earth. The valve is assumed to be non-conducting until it is put into operation by the leading edge of a rectangular pulse applied to the grid, so that initially $v_K = 0$ and v_i is below cut-off; (Fig. 599).

As v_i rises through cut-off valve current begins to flow and the rise of v_K follows that of v_i in a manner similar to the action of a cathode follower (see Chapter 7 Section 18), the cathode load being equal to the characteristic impedance of the network. The initial rise of v_K is slightly less than the amount by which v_i rises above cut-off. The voltage change v_K travels along the network to the open-circuited end, where it is reflected, without a change of sign, to return back along the network. The returning voltage change arrives between cathode and earth an interval $2n\sqrt{LC}$ after the start of the operation, and

increases the voltage of the cathode by an amount sufficient to cut off the valve current. The leading edge of the output pulse, appearing across the anode resistor R_f , is produced as the rise of input voltage causes the valve to conduct. The trailing edge of the output pulse is formed as the action of the network causes the valve current to be cut off. Once the valve current is cut off the cathode voltage falls slowly as the network discharges through the terminating resistor R_f . The time needed for this discharge should be long enough to ensure that the valve current remains cut off until the return of the input voltage to its initial value.

High voltage pulses of about 1 microsecond duration, such as are required for the modulation of transmitters, can be obtained by first charging a delay line to a high voltage and then discharging it through a resistive load. A detailed consideration of this method of pulse production is given in Chapter 14.

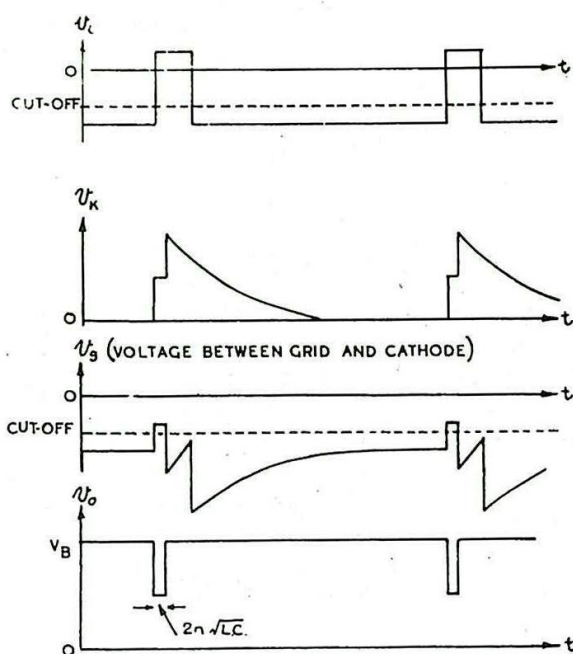


Fig.599.- Response of delay line in cathode circuit.

CHAPTER 14

PULSE MODULATION OF TRANSMITTERS

1. MODULATION REQUIREMENTS

In Chap. 8 various types of oscillators are described, mainly with regard to their ability to generate continuous oscillations. In pulse radar systems, power oscillators are not employed to generate CW but RF pulses. However, since the time taken for oscillations to build up is much less than the duration of the output pulses in present-day Service equipments, this difference does not invalidate the description of the modes of oscillation given in Chap. 8.

Transmitting valves are usually held quiescent for comparatively long periods, of the order of a millisecond. They are then caused to operate, generating a RF pulse lasting for about one microsecond, after which they remain inoperative until the next pulse is required. The unit which controls the Firing or Pulsing of the oscillator is called the Modulator. This term is retained from Radio Communication, in which the RF output is modulated in a manner which conveys the intelligence. It is usually not convenient to regard the modulator from the point of view of the integrated effect of recurrent pulses but to consider it instead as a switching device, controlling the initiation, and, sometimes, the shape and the duration of the output pulses. Considered in this manner it is the function of the modulator to maintain the oscillator valve quiescent during the intervals set aside for reception, and to bring it into oscillation as and when required, maintaining the controlling potentials at suitable values throughout the duration of the pulse. Also, except in the case of self-quenching oscillators, the modulator controls the duration of the output pulses.

In some systems the modulator controls the recurrence period by its own independent timing mechanism. This may be necessary when spark gap modulators are employed, since these are prone to jitter so that they cannot readily be triggered to spark at an exact instant, and the uncertainty may lead to unnecessary timing errors. Either the modulator pulse or a rectified version of the transmitter pulse may be used to start the timing mechanism for each successive recurrence period. Where gas-filled triodes or hard valves are used as modulator valves, or when the errors due to spark-gap jitter are not important, the modulator may be externally controlled by a master timing system.

In pulse radar as commonly employed successive transmitter pulses are not phase-linked and are said to be Non-Coherent. Some types of radar system require Coherent pulses; i.e. successive pulses must occur at regular intervals of time and with a regular phase relationship. In such systems the problem of modulation is more complex. In this chapter we shall consider modulators for the production of non-coherent RF pulses only.

2. MODULATION METHODS

Pulse modulators may first be divided into two main types:-

I .Grid Modulators

II Anode Modulators

Although many early radar systems employed I this is now generally abandoned in favour of II for the following reasons.

- (a) It is necessary with grid modulation to maintain a high HT voltage for much longer than with Type II. This wastes power and necessitates the use of larger and more elaborate equipment.
- (b) More precise control is possible with anode modulation.
- (c) Most transmitter valves (magnetrons) now used for operation at centimetre wavelengths have no control grids.

Anode modulators may be classified under two general headings:-

- (i) Soft Valve Modulators.
- (ii) Hard Valve Modulators.

This division is more profound than the mere names suggest. Because of the high power handled by the anode modulator, which must apply to the oscillator valve the necessary HT voltage (usually several kilovolts), and which, during the pulse, carries the full current of the oscillator valve (usually several tens of amps) it is a decided advantage to use soft valves, which are capable of passing larger currents with less dissipation than hard valves. Unfortunately, soft valves, although readily triggered, are not so readily extinguished. It is necessary, when using a soft modulator valve, to employ a pulse-forming network as a supply source for the oscillator valve. The power available for each pulse is then utilised during the double transit time of the network and the oscillator is therefore unable to oscillate after that time.

Where hard valve modulators are used there is no such limitation. The shape and duration of the modulator pulse are determined by a pulse forming network at a low power level. Subsequently the pulse is amplified by hard valve amplifiers which are controllable, both at the beginning and at the end of the pulse, by the potentials of their control grids.

The divisions (i) and (ii) have, therefore, the following significance.

(i) The pulse-forming network is made to carry the whole of the pulse energy and is discharged into the oscillator in series with the soft modulator valve, which controls the start of the discharge. The problem of charging the network and causing it to discharge as and when required is the main consideration.

(ii) The pulse forming network is incorporated in a low-power stage and the pulse is subsequently amplified until the power level is sufficient for modulating the oscillator. The problem of amplification, and the choice of suitable valves, is the main consideration.

Both systems are commonly used in Service equipments.

GRID MODULATION

3. Pulse Amplifier Method

The essential circuit of this method is shown in Fig. 600.(a).

The oscillator is shown for simplicity as a single valve, valve 1, with lumped L-C tuning, although arrangements using push-pull valves and tuned lines may be employed. The RF connection from the tuned circuit back to earth is via the condenser C, but the DC path is via the resistor R, which forms the anode load of the pulse amplifier, valve 2. The cathode of valve 2 is returned to the negative supply line, $-2V_B$. The action is as follows.

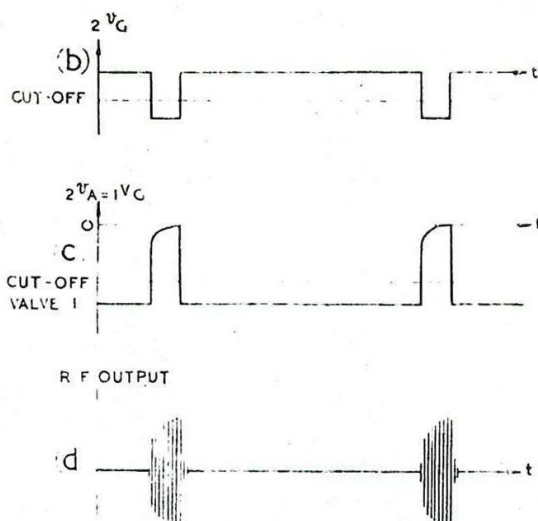
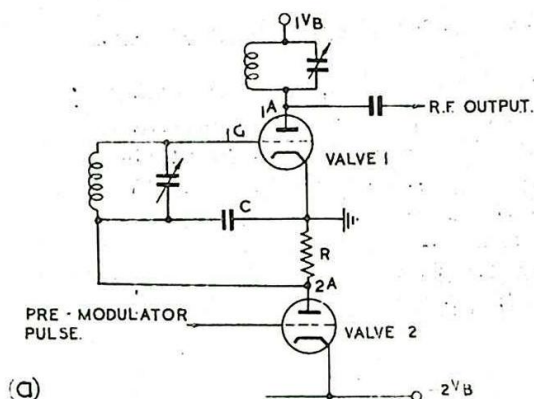


Fig. 600 - Grid modulation: pulse amplifier method.

- (i) Valve 2 is normally conducting heavily, since its grid-cathode potential is approximately equal to zero. The flow of anode current through R depresses the potential at 2A, and therefore the grid bias of valve 1, below earth by an amount sufficient to cut off the anode current of valve 1, which therefore does not oscillate.

- (ii) The anode current of valve 2 is cut off by means of a negative pulse applied to its grid, so that its anode voltage, and hence the grid bias of valve 1, rises towards earth potential. This allows valve 1 to conduct and generate oscillations. The waveforms are shown in Fig. 600.

Capacitance in shunt with R, including the anode-cathode capacitance of valve 2 and that of the RF bypass condenser C, delays the rise and fall of the voltage at 2A, and the grid bias of valve 1. The effect is most marked on the leading edge of the pulse, which rises with a time-constant equal to the product of the total shunt capacitance and the load R; in order to make this time-constant small, R must have a small value (e.g., $3k\Omega$). A second reason for making R small is that the grid current of the oscillator flows through it, tending to produce a negative bias. To obtain the required rise in voltage at 2A (e.g. 1500 volts) across a low resistance requires a large decrease of current (e.g. 0.5 amp). Thus valve 2 must be capable of passing such a current continuously and have an adequate permissible anode dissipation, because it is conducting for the whole of the recurrence period except for the duration of the RF pulse. For this reason, the mean power taken from the HT supply $-2V_B$ is large. A suitable value for $2V_B$ is 2.5 kV.

4. Cathode Follower Method

Fig. 601(a) shows the circuit for this method. The DC return path from the grid circuit of the oscillator (valve 1) is via the cathode load R of the cathode follower (valve 2). The operation is as follows:-

- (i) Valve 2 normally biases itself nearly to cut-off, because its grid circuit is returned to the negative end of R , the voltage at $2K$ becoming positive with respect to $-2V_B$ by an amount approaching the cut-off bias of valve 2. This means that the voltage at $2K$, and hence the grid bias of valve 1, are sufficiently negative with respect to earth to cut off the anode current of valve 1.

- (ii) A positive pulse applied to the grid of valve 2 raises the grid voltage to a value approaching earth potential, and the cathode voltage, together with the grid bias of valve 1, follows, so bringing valve 1 into conduction.

Capacitance in shunt with R produces little delay in the rise of the pulse because of the low output impedance of the cathode follower. The trailing edge, however, may be adversely affected, since for this the output resistance of the cathode follower is larger than for the leading edge. The fall in voltage at $2K$ tends to lag behind that at $2G$ so that the valve current is cut off. The output resistance for the trailing edge of the pulse is thus approximately R , and the corresponding time-constant is the product of R and the total shunt capacitance.

The waveforms for the cathode follower circuit are shown in Fig. 601.

Advantages of this method compared with the former are:-

- (i) The modulator valve, valve 2, is conducting heavily only during the pulses; consequently its anode dissipation during the quiescent period is comparatively low and the

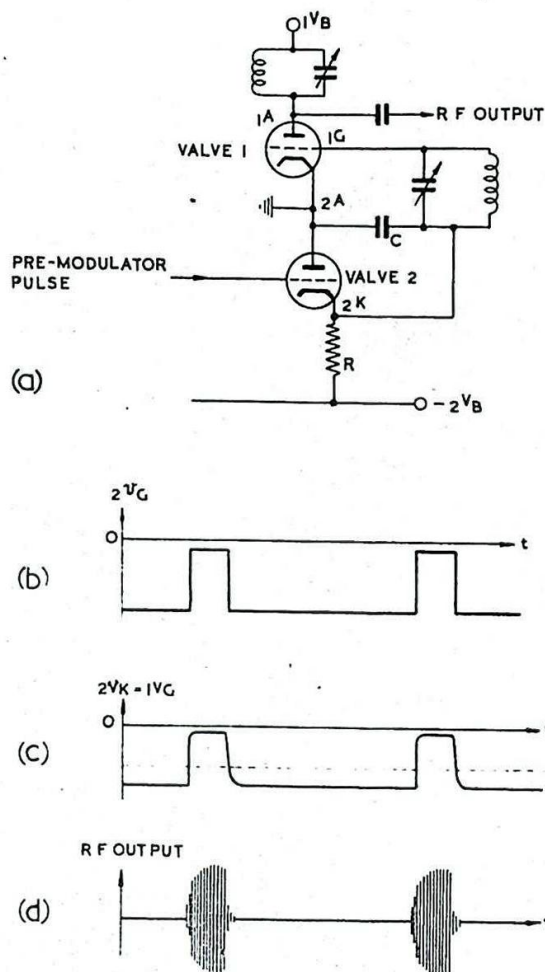


Fig. 601 - Grid modulation: cathode follower method.

valve is operated more efficiently. However, the cathode of valve 2 must still have a high emissivity because, while valve 2 is conducting, it must carry the grid current of valve 1.

- (ii) The distortion of the pulse due to shunt capacitance is reduced.

The cathode-follower modulator requires to be driven by a positive-going pulse of amplitude somewhat greater than that of the output pulse required. The stage generating this pulse can be an amplifier of the type described in Sec. 3 but can have a larger output resistance than is permissible if such a stage is employed as the modulator. This is because the input impedance of the cathode follower is considerably higher than that presented by the grid circuit of the oscillator.

The design of a grid modulator is similar to that of the Pre-Modulator Stage for a hard valve anode modulator. It is therefore suggested that Secs. 9 and 10 should be read in conjunction with the above.

5. Grid Self-Quenching

If an oscillator is provided with a C-R network in its grid circuit, it may be made to generate RF pulses. The phenomenon of Grid Self-Quenching (squegging) has already been described in Chap. 8, Sec. 47. It is sufficient here to indicate the circuit to be used and the waveforms produced, illustrated in Fig. 602. In Fig. 602(a) the condenser C is shown inserted between the top end of the L-C circuit and the grid, but if more convenient it may be placed between the bottom end and earth. The resistor R is connected either to zero or to a positive potential.

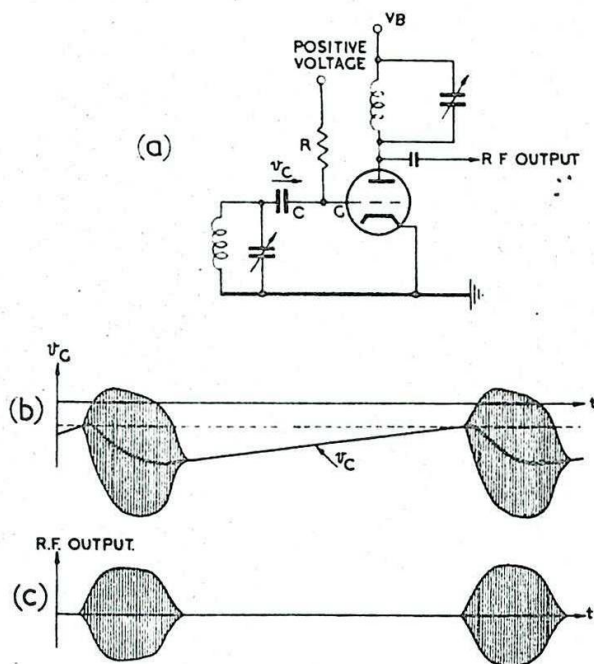


Fig. 602 - Grid self-quenching oscillator.

The squegging oscillator has the merit of extreme simplicity but possesses several disadvantages:-

- (i) The bias is varying throughout the duration of the pulse so that optimum conditions of oscillation are not maintained.
- (ii) The pulse shape is not very good.
- (iii) The repetition period is rather indeterminate because it depends on the slow exponential rise of the grid voltage to cut-off. Variation of cut-off potential with ageing of the valve or between different valve specimens can thus cause alternations of recurrence

frequency. This tendency is reduced when R is returned to a positive voltage.

(iv) The RF pulse length is also rather indeterminate because it depends on the rapidity with which C charges due to grid current flow. Change of grid-current characteristics with ageing of the valve or change of valve specimen thus produces variation of pulse width.

The indeterminacy of the recurrence frequency of the freely running squegging oscillator can be eliminated by triggering the grid circuit of the oscillator by one of the methods described in Secs. 3 and 4. The triggering pulse is longer than the required duration of the RF pulse, the latter being determined as before by the rate at which C is charged by grid-current flow. The circuit may be arranged as in Fig. 603, a cathode follower triggering valve being used. From the waveforms it can be seen that the action of the oscillator is initiated by the triggering pulse and its duration is determined by the squegging action. The free-running repetition frequency is chosen to be greater than the frequency of the triggering pulses so that the valve is ready for operation by the time each triggering pulse is applied.

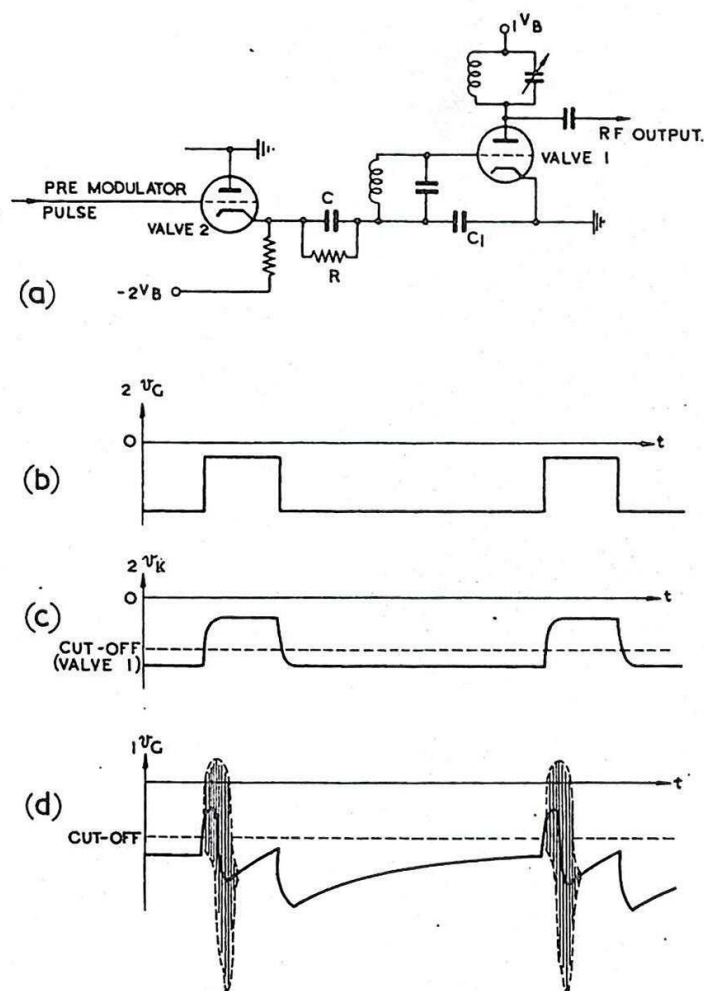


Fig. 603 - Triggered squegging oscillator.

The duration of the trigger pulse need not be shorter than is necessary to prevent double pulsing. Double pulsing occurs if the rate of rise of mean grid potential after the end of the RF pulse is such that cut-off is reached before the end of the positive portion of the triggering

pulse. In this case a second RF pulse is produced by the oscillator.

The advantage of the arrangement compared with the simple cathode follower type of grid modulator is that it removes the necessity of generating triggering pulses as short as the duration of the RF pulse.

6. Anode Self-Quenching

Although this is not a method of obtaining grid modulation, it is convenient to mention it here, as it is used in conjunction with triggered grid self-quenching circuits to reduce the likelihood of double pulsing. If the anode circuit of the oscillator is fed from the HT supply via a C-R network as shown in Fig. 604, the applied HT, instead of remaining constant as with direct feed, varies in the manner shown in Fig. 604. The applied HT falls rapidly due to the discharge of C by anode current flow while the valve is oscillating during the first part of the trigger pulse; it then commences to rise as C recharges through R, but the rise during the remainder of the trigger pulse is small. Because of the low value of applied HT during this period, the cut-off value of grid potential is less negative; consequently the mean grid potential can rise further during the remainder of the triggering pulse without the oscillator coming into conduction and producing a second RF pulse. The variation of applied HT also affects the RF pulse length and shape.

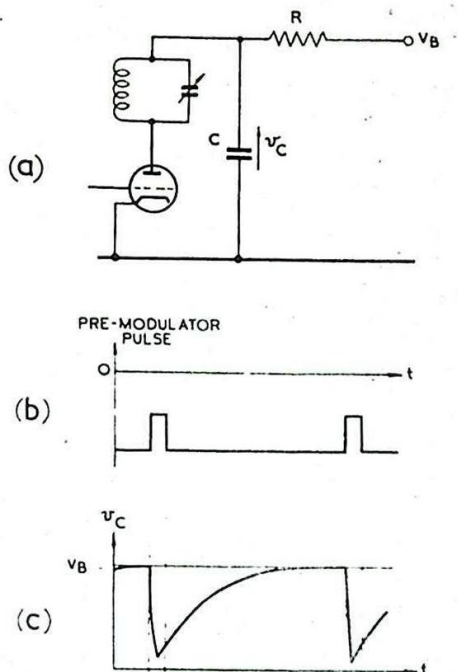


Fig. 604 - Anode self-squegging.

The action of the triggered self-quenching circuits can be made more definite by substituting an open-ended delay network for the condenser C of the C-R circuit. The operation of self-quenching using delay networks will not be described as they have not been used to any great extent. Anode self-quenching using a delay network constitutes one of the important methods of anode modulation and is described in Sec. 11.

7. Disadvantages of Grid Modulation

(i) If valves with oxide-coated cathodes are used grid modulation may be seriously affected because of grid emission caused by deposition of oxide-coating on the grid. Gold plating the grid reduces this effect, the gold poisoning the sputtered coating and preventing emission. Even with thoriated tungsten filaments grid emission can occur for either of two reasons.

(a) If the grid, by reason of its proximity to the cathode, becomes very hot, primary emission from the grid may occur.

(b) If the grid is driven very positive, secondary emission may result.

In any case the Reverse Grid Current tends to cause the squegging network to charge up in the wrong sense, with possibilities of CW oscillations ensuing. In general, special precautions must be taken to prevent such reverse current from passing through the squegging network, such as the use of bypass diodes.

(ii) The HT is applied to the oscillator continuously, so increasing the tendency for the phenomenon known as flash-arcing to occur; this is the complete breakdown of the high insulation normally afforded by the high vacuum between the electrodes of the valve.

ANODE MODULATION

8. General

In this type of modulation, the mean potential of the grid (if any) of the oscillator remains constant, whilst the anode-cathode voltage is switched to the operating value for the required duration of the RF pulse. This action may be brought about by switching either anode or cathode voltage, the latter being used for oscillator valves which work with their anodes earthed.

9. Hard Valve Modulator

The basic circuit for Hard Valve Modulation is shown in Fig. 605. The oscillator and modulator valves are connected in series across the HT supply. The modulator-valve current is normally cut off by the negative grid bias and no current can flow in the oscillator valve. A positive pulse is supplied to the modulator grid so as to bring the valve into heavy conduction for the required duration of the RF pulse. During the pulse, the modulator valve should offer a low impedance so that the greater part of the HT voltage may be developed across the oscillator valve.

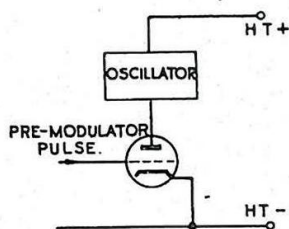


Fig. 605 - Basic circuit of hard valve modulator.

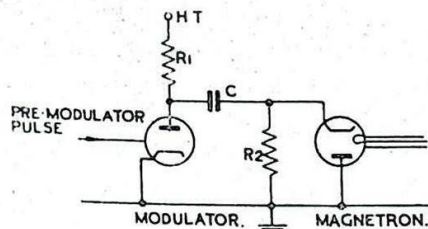


Fig. 606 - Shunt-fed modulator circuit.

If a magnetron oscillator is employed the HT positive lead must be earthed since the anode of the magnetron is at earth potential because of its mechanical construction. This may be avoided by adopting a shunt-fed arrangement, as in Fig. 606. To ensure correct operation the voltage developed across the magnetron must remain steady during the pulse. This means that a long time-constant coupling circuit should be used. The resistances involved in this coupling circuit are:-

R_1 and R_2

R_v , the resistance of the modulator valve when conducting, and

R_m , the resistance of the magnetron when oscillating.

For maximum efficiency the following results must hold:-

$$R_1 \gg R_v$$

$$R_2 \gg R_m$$

$$R_m \gg R_v$$

Hence the condition for small distortion of the rectangular pulse is that $CR_M \gg$ the duration of the pulse.

10. Hard Modulator Valves

The Modulator valve must satisfy the following requirements:-

- (i) The cathode must have a high emission, since the oscillator valve current passes through the modulator valve.
- (ii) The resistance of the valve when conducting must be small compared with that of the oscillator in order that the greater part of the HT voltage may be developed across the oscillator.
- (iii) The valve must be capable of withstanding the full HT.
- (iv) The valve must have small inter-electrode capacitances in order to preserve pulse shape.
- (v) The maximum permissible anode dissipation must be adequate.

These different requirements conflict with one another, and the design is necessarily a compromise.

Two examples of modulator valves are given below :-

(i) CV57 Beamed Tetrode

Thoriated cathode. Gold-plated grid.
Peak emission = 5A
Impedance = 500 Ω
Maximum anode voltage = 11 kV
Input capacitance = 21 pF
Output capacitance = 11 pF
Maximum anode dissipation = 15W

Three of these valves in parallel are normally used for modulating a CV64 magnetron.

(ii) 715C Beamed Tetrode

Oxide-coated cathode. Gold-plated grid
Peak emission = 15A
Impedance = 300 Ω
Maximum anode voltage = 20kV
Input capacitance = 35pF
Output capacitance = 8 pF
Maximum anode dissipation = 60W

11. Anode Modulation using a Delay Network in Conjunction with a Spark Gap or a Gas Triode

In this method the modulating pulse is obtained by discharging a delay network into the load presented by the oscillator through a spark gap or a gas triode. The delay network is charged from the HT source in the intervals between pulses.

The principle of operation may be understood by first

considering the action of the circuit of Fig. 607(a). This consists of the anode self-quenching circuit of Fig. 604, with the addition of a switch S in series with the oscillator. In practice, S takes the form of a spark gap or a gas triode; these possess the property of maintaining conduction at very much lower voltages than those at which they strike.

While S is open, the condenser C charges through resistor R to the HT voltage. When S is closed, C discharges through the load presented by the oscillator, the resistance of R being very much greater than that of the load. When the voltage across C reaches some value V_1 , Fig. 607(b), S opens (i.e. the spark gap or gas triode extinguishes) and C recharges through R . The voltage across the oscillator is zero between pulses and, neglecting the small loss of voltage across S , is equal to v_c while S is closed.

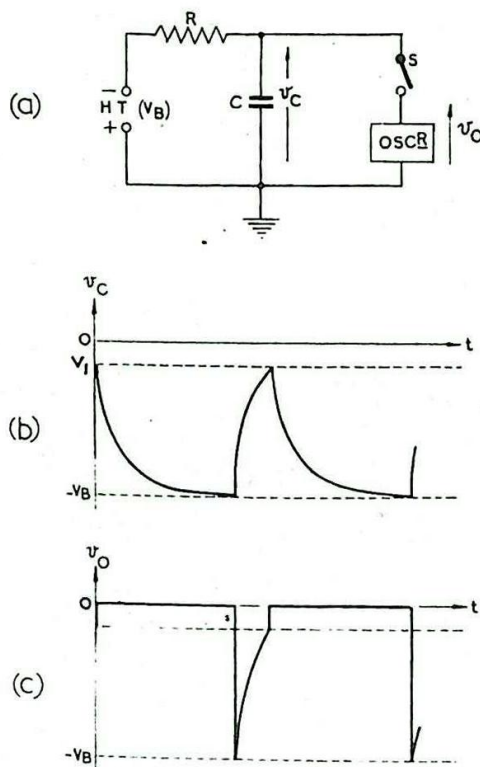


Fig. 607 - Anode modulation using C-R network.

The circuit just described is capable of generating modulating pulses for the oscillator, but the pulse shape is far from rectangular. This disadvantage may be overcome by replacing C by an open-ended delay network as shown in Fig. 608. While S is open, the condensers charge through R as if they were all connected in parallel, the inductances having negligible effect during the relatively slow charge. The total charging capacitance is thus nC ; (b). When S is closed the network discharges producing a more or less rectangular pulse across the oscillator; (c). The manner of this discharge depends on the following considerations:-

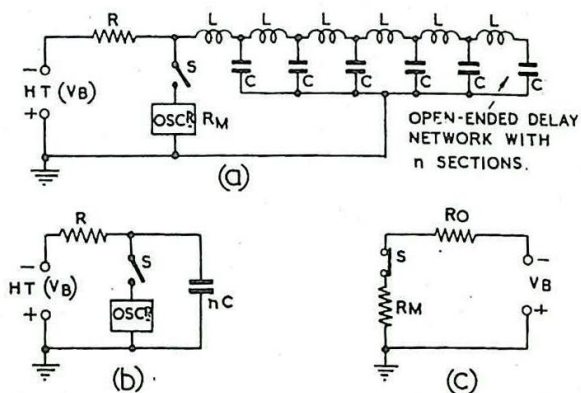


Fig. 608 - Anode modulation using delay network.

- (i) The number of sections in the delay network. The more sections there are the more nearly rectangular the pulse becomes.
- (ii) The value of LC for each section. The duration of the

pulse is $2\pi\sqrt{LC}$.

- (iii) The ratio of the characteristic impedance of the network at zero frequency to the resistance of the load. The existence of subsequent reflections, their magnitude, and polarity, depend on this ratio.

These effects are dealt with in Chap. 4, Sec. 12 and Chap. 3, Sec. 16. Various methods of charging the delay network may be employed. These are dealt with in Secs. 23 - 36.

Owing to the high voltages used, the network is usually immersed in oil to minimise leakage and the possibility of a flash-over.

PULSE TRANSFORMERS

12. General

High voltage transformers normally used in power supply circuits are designed to operate with currents which are approximately sinusoidal, and are required to produce a corresponding sinusoidal output voltage. In high-power pulse circuits the requirements are different, and ordinary transformers are liable to distort pulse-shape very considerably.

The term Pulse Transformer is applied to transformers which are designed to pass pulses without appreciable distortion of their shape.

Pulse transformers find particular application in the modulator section of radar equipments, being used for the following purposes:-

- (i) Matching, e.g. matching the modulator or magnetron to the cable connecting them.
- (ii) Phase-reversal of a pulse, e.g. between the pre-modulator and modulator valves.
- (iii) Coupling between two circuits at different steady voltage levels.

13. Equivalent Circuit of a Pulse Transformer

The distortion of the pulse shape produced by a pulse transformer can be determined by analysis of its equivalent circuit. An approximate equivalent is shown in Fig. 609. A mathematical justification for this equivalent circuit can be given, but it is sufficient here to indicate the significance of the various circuit elements shown.

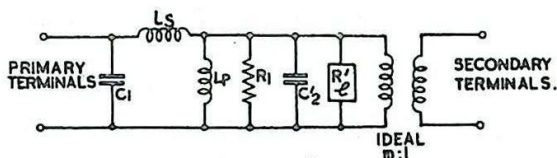


Fig. 609 - Equivalent circuit for pulse transformer.

C_1 is the primary winding self-capacitance + output capacitance of the stage feeding the primary.

L_s represents the leakage inductance between primary and secondary i.e. the inductance of the primary when the secondary is short-circuited.

L_p represents the magnetising inductance, i.e. the inductance of the primary when the secondary is open-circuited.

R_1 represents the losses due to dissipation in the iron core.

m is the step-down ratio of the transformer (the turns ratio).

$C_2' = C_2/m^2$, where C_2 is the secondary winding self-capacitance + stray shunt capacitance + input capacitance of secondary load.

$R_1' = m^2 R_1$, where

R_1 is the secondary load resistor.

C_2' and R_1' are the equivalent values of C_2 and R_1 reflected from the secondary to the primary circuit of the transformer.

The Ideal Transformer merely transforms the voltage in the ratio of the turns, and introduces no other circuit elements.

14. Pulse Transformer Connected Between a Modulator Valve and Magnetron

An approximate treatment of the operation of a pulse transformer connected between a modulator and magnetron (Fig. 610(a)) is now given. The modulator and magnetron valves conduct heavily during the pulse, and are non-conducting between pulses. The equivalent circuits for the duration of the pulse, and for the interval between pulses are considered separately.

(i) Operation during pulse

The equivalent circuit during the pulse may be derived from Fig. 609 and is shown in Fig. 610(b). In this figure

R_v = anode-cathode resistance of the modulator valve when conducting.

$R_M' = m^2 R_M$ where

R_M = anode-cathode resistance of magnetron when conducting.

V_M = voltage across magnetron.

R_1 of Fig. 609 is omitted because $R_1 \gg R_M$.

This still complicated equivalent circuit can be simplified by considering separately the action of the circuit for the leading edge of the pulse (modulator just switched on) and during the steady portion of the pulse.

(a) Response on leading edge.

The equivalent circuit may be simplified by omitting C ,

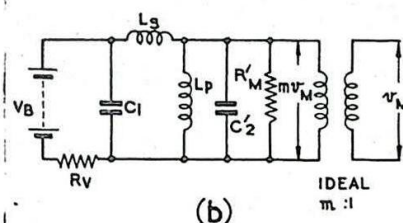
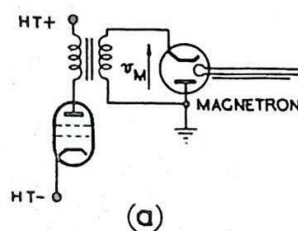


Fig. 610 - Pulse transformer connected between modulator valve and magnetron.

since C can charge very rapidly through the low impedance R of the modulator. We may also omit L_p for the following reasons:-

In the first instance neglect L_s since $L_s \ll L_p$. Then the current through L_p is small provided the time-constant $\frac{L_p}{R_v}$ is much greater than the corresponding time-constant for the circuit formed by C_2 , R_M and R_v , i.e.

$$\frac{L_p}{R_v} \gg \frac{C_2' R_M' R_v}{R_M' + R_v}$$

Normally $R_M' \gg R_v$ so that this reduces $\frac{L_p}{R_v} \gg C_2' R_v$

This condition is normally satisfied in a pulse-transformer circuit. The simplified circuit is shown in Fig. 611(a); in this the ideal transformer is omitted for clarity.

The voltage v_M across the magnetron does not instantaneously attain its steady value of $-\frac{V_B R_M'}{m(R_M' + R_v)}$, because of the presence

of L_s' and C_2' which form a series resonant circuit. Oscillations occur if the damping factor ζ of the circuit is less than unity; (see Chap. 2, Sec. 10). Any reduction in damping below this critical value will

- (1) introduce a ring, which is undesirable;
- (2) steepen the leading edge and so accelerate the build-up, which is desirable.

A compromise must be achieved between these effects, and the value $\zeta = 0.7$ is commonly used. This compromise is illustrated in the following typical specification:- "The pulse must attain 90% of its final voltage within 0.1 μ sec, and the amplitude of the ripple must not exceed 10% of the mean pulse voltage." The pulse shape for various values of ζ is shown in Fig. 611(b).

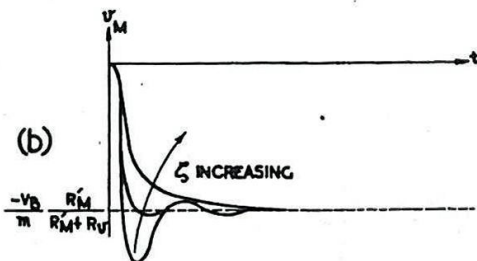
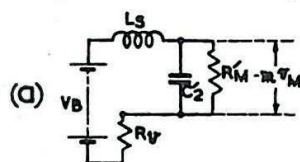


Fig. 611 - Performance of pulse transformer during leading edge of pulse.

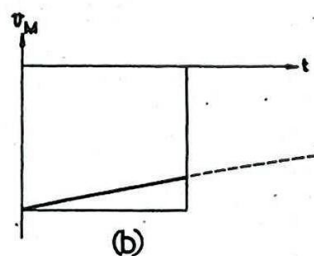
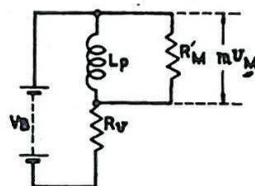


Fig. 612 - Performance of pulse transformer during steady portion of pulse.

(b) Response during steady portion of pulse.

Here the gradual build-up of magnetising current in L_p results in a decrease of current through the magnetron. Since this variation of current is relatively slow, C_1 , L_s , and C_2' may all be neglected. The resultant simplified circuit is shown in Fig. 612(a).

The voltage mV_M across L_p falls from its initial value with a time constant

$$\frac{L_p (R_M' + R_V)}{R_V R_M'} \approx \frac{L_p}{R_V}$$

the pulse shape being as depicted in Fig. 612(b).

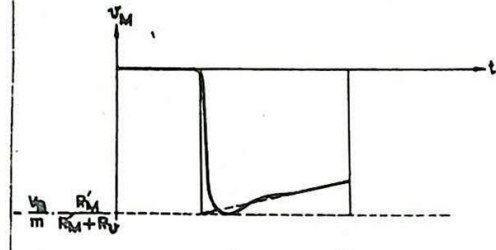
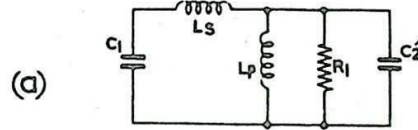


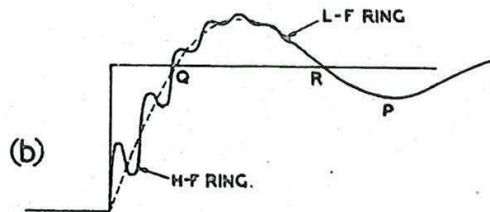
Fig. 613 - Pulse transformer: combined effects on leading edge and steady portion of pulse.

(ii) Operation between Pulses

We will consider the effects on the trailing edge of the pulse, when the modulator and magnetron have become non-conducting. The equivalent circuit is shown in Fig. 614(a). C_1 and C_2' are charged and must discharge before equilibrium is reached. Two types of ring can occur:-



- (a) A relatively low-frequency ring due to L_p resonating with the parallel combination of C_1 and C_2' , L_s having little effect for this slow oscillation. Damping occurs due to R_1 .



- (b) A high-frequency ring due to L_s resonating with the series combination of C_1 and C_2' , L_p having little effect on this rapid oscillation.

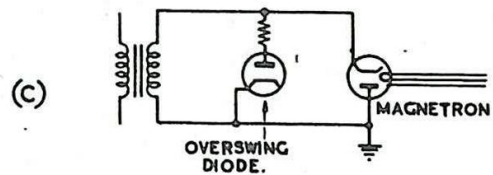


Fig. 614 - Performance of pulse transformer after the end of the pulse.

Fig. 614(b) shows the type of waveform produced, the high-frequency and low-frequency rings being superimposed.

15. Design Considerations

It follows from Sec. 14 (i) that for minimum distortion during the leading edge and steady portion of the pulse the constants must be chosen so that

$$\frac{L_p}{R_v} \gg T_p \gg C_2' R_v,$$

where T_p is the duration of the pulse.

Consequently C_2' should be as small as possible, with a large value of L_p .

In order that the trailing edge of the pulse should be sharp, the frequency of the low-frequency ring should be high; i.e.

$L_p (C_1 + C_2')$ should be small.

In general these two sets of conditions cannot be attained unless R_v is very small. A large value of L_p can be obtained by either using a large number of primary turns or using a high permeability core. Of these, the latter is preferable because smaller values of L_s , C_1 and C_2 are then obtained. High-permeability core materials are available in the form of certain nickel-iron alloys, of which the outstanding example is mu-metal. A considerable reduction of leakage inductance can be effected by the use of an auto-transformer, rather than one with two separate windings. This is possible only when primary and secondary circuits are at the same steady potential.

The insulation of high voltage pulse transformers is a difficult problem; the most suitable method is to enclose the windings in a sealed oil-filled container. Allowance must be made for expansion of the oil due to heat.

16. Overswing Diode

In Fig. 614(b), the negative peak at P may reach sufficient amplitude to cause the magnetron to oscillate again, leading to the production of a double RF pulse. To prevent this a diode is connected, as in Fig. 614(c) so that it conducts on the preceding positive half-cycle, i.e. between Q and R, and damps out the oscillation. A protective resistance is usually included in series with the overswing diode.

17. Pulse Transformer with Bifilar Secondary

If a pulse transformer is used to feed the HT pulse to the cathode of the magnetron (Fig. 615) the transformer supplying the heater of the magnetron must be rated to withstand the full pulse voltage. This may be avoided by the use of a pulse transformer with a bifilar secondary, connected as in Fig. 616. The two sections of the bifilar winding are numbered 3, 4 and 5, 6. The ends marked 4, 6 are effectively at earth potential so far as the pulse is concerned, 6 being connected to earth directly and 4 being connected thereto via the condenser C_1 (e.g. 0.01 μ F). The heater voltage for the magnetron can therefore be applied at the points 4, 6 from a transformer of normal rating.

Equal pulse voltages appear at the ends of the winding of the pulse transformer marked 3, 5 and are applied to the two sides of the heater of the magnetron, one of which is connected to the cathode. The condenser ensures that any inequality in the pulse outputs of the bifilar winding does not damage the heater.

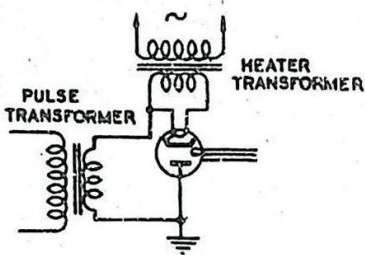


Fig. 615 - Pulse transformer feeding magnetron.

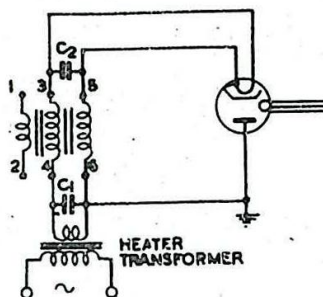


Fig. 616 - Pulse transformer with bifilar secondary.

18. Use of Pulse Transformer in Conjunction with Hard Valve Modulators.

In certain Service applications it is necessary to separate the oscillator and modulator valve circuits. To avoid distortion of the pulse shape by the connecting cable it is necessary to terminate the cable in its characteristic impedance; this may be done by using a pulse transformer to match the impedance of the magnetron into the cable. The matched cable will not in general present a suitable impedance for direct connection to the modulator, and a second pulse transformer is inserted between the modulator valve and the input of the cable. This arrangement is shown in Fig. 617; since the impedance of the cable is normally considerably less than that of the magnetron, the input pulse transformer has a step-down turns ratio, and that at the output a step-up ratio. Because of this, the voltage across the cable is considerably less than that applied to the magnetron, with the consequence that the necessary voltage rating of the cable is much reduced.

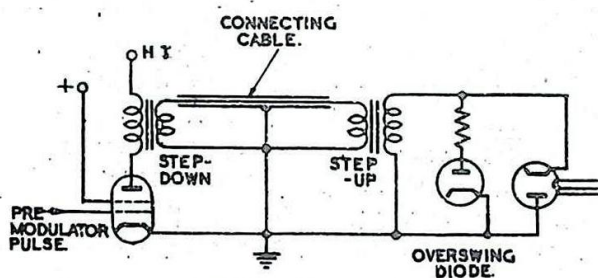


Fig. 617 - Use of connecting cable and pulse transformers between modulator and oscillator.

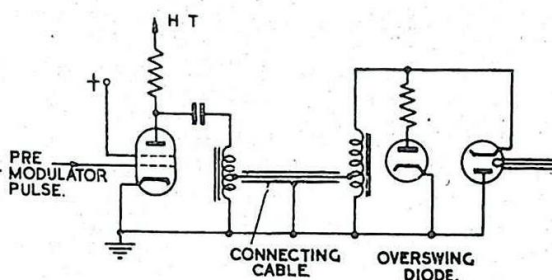


Fig. 618 - Use of autotransformers between modulator and oscillator.

Example:-

Impedance of magnetron when conducting = $900\ \Omega$

Characteristic impedance of cable = $100\ \Omega$

Amplitude of pulse across magnetron = $12\ \text{kV}$.

Step-up turns ratio required between cable and magnetron

$$\frac{900}{100} = 3.$$

Voltage across cable = $\frac{12}{3}\ \text{kV} = 4\ \text{kV}$.

The purpose of the overswing diode in Fig. 617 is described in Sec. 16.

In place of the double-wound transformer shown (Fig. 617) auto-transformers are frequently employed. An arrangement using auto-transformers is drawn in Fig. 618.

19. Use of Pulse Transformer in Conjunction with Delay Network

(Soft Valve) Modulators

Pulse transformers may be used in conjunction with a delay network modulator for the following purposes:-

(i) To step up the voltage of the pulse before application to the oscillator. This allows the use of a delay network with a lower voltage rating. It increases, however, the current which must flow through the spark gap or gas triode to produce a given current in the oscillator circuit, e.g. if 12kV. 10A. are the voltage and current required in the oscillator circuit, and 4kV the pulse voltage across the delay network, a current of 30A. must flow through the switching device.

(ii) To enable a low impedance cable to be used between modulator and oscillator units.

HARD VALVE MODULATOR CIRCUITS

20. Pre-Modulator

A hard modulator valve requires the application of a positive-going pulse of large amplitude (e.g. 1,000 V) to bring it into conduction. The unit which generates the pulse is known as the Pre-Modulator (or Driver).

The action of the pre-modulator is initiated by a triggering pulse from the radar timing mechanism. This pulse is normally of comparatively long duration and is used to produce a short duration pulse of the desired shape by one of the methods described in Chap. 13. This pulse may go through several stages of amplification, inversion, shaping, etc. before the output stage is reached.

The output impedance of the pre-modulator unit must be low, for the following reasons:-

(i) The grid of a large modulator valve must be driven considerably positive, e.g. 100 V, in order to make the valve conduct as heavily as possible; under these conditions the input impedance is low due to the flow of grid current. Unless the output impedance of the pre-modulator is sufficiently low, limiting will occur as described in Chap. 9.

(ii) A low output impedance is necessary to preserve pulse shape.

The simplest form of pre-modulator output stage consists of a valve with resistive anode load which conducts during the intervals between pulses and is rendered non-conducting by the application of a negative pulse to its grid for the required duration of the RF pulse. In order that the stage may have a low output impedance, it is necessary to use a small anode load. To produce a voltage pulse of large amplitude across a small resistance requires a large current change. Therefore a valve capable of passing a heavy current continuously and having a large permissible anode dissipation must be used,

On the other hand the stage supplying the negative pulse which cuts off the current of the output stage can be of low power since it is required to conduct only for the duration of the RF pulse.

The high power consumption of the pre-modulator output stage may be avoided by inserting a phase-reversing pulse transformer between the anode of the output stage and the grid of the modulator valve. In this case the output valve is brought into conduction for the duration of the RF pulse only, and although it must be capable of handling large peak currents, its ratings of maximum emission current and anode dissipation may be comparatively low. Such a stage may, if desired, be driven from a normally quiescent low-power stage through another phase-reversing transformer.

A representative arrangement for a complete pre-modulator unit is shown in Fig. 619 together with the waveforms of the voltages produced at various points. The negative-going edge of the rectangular triggering voltage corresponds to the firing instant of the transmitter; this voltage is applied to the grid of valve 1 via a long

time-constant circuit, which in conjunction with the grid-cathode circuit of the valve, clamps the voltage so that its positive-going level at the grid is approximately zero volts. In the anode circuit of valve 1 is a short-circuited delay network, which results in the production of positive-going and negative-going pulses of the required duration. These are applied to the grid of valve 2 which is biased so as to be brought into conduction only during the positive-going pulses. The amplified negative-going pulses at the anode of valve 2 are reversed in phase by the pulse transformer T_1 and applied to the grid of valve 3, which is also biased so as to conduct only during the (positive) applied pulses. The negative-going pulse output of valve 3 is reversed in phase by T_2 before application to the modulator valve.

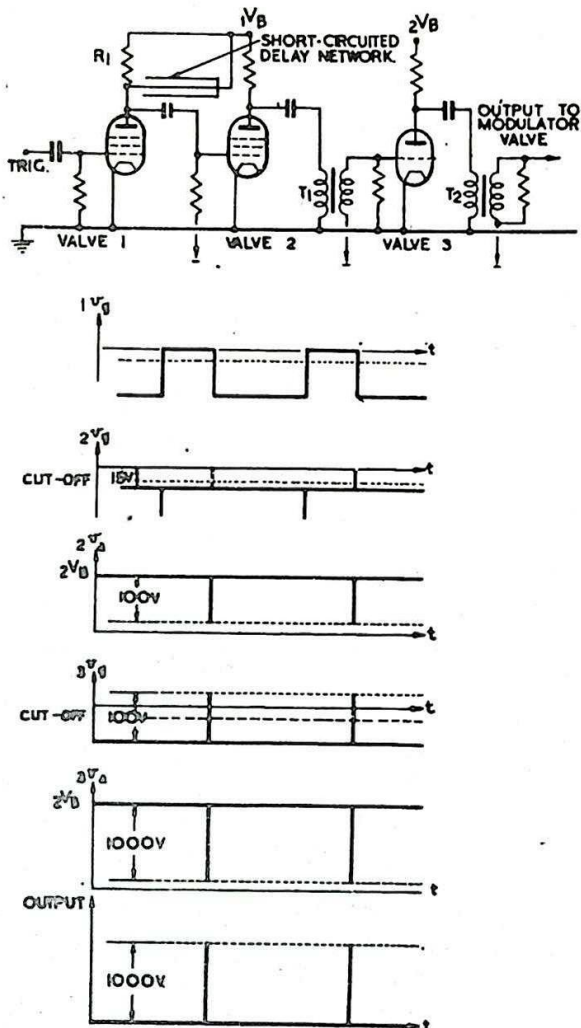


Fig. 619 - Representative pre-modulator circuit and waveforms.

21. Boot-Strap Pre-Modulator

The employment of pulse transformers, with the consequent possibilities of distortion of the pulse shape, can be avoided, whilst

at the same time all valves may be kept in the quiescent state except during the pulse, by use of the Boot-Strap circuit. This circuit, together with waveforms, is shown in Fig. 620.

In the interval between pulses valve 2, a gas triode, is biased by the voltage $-V_G$ so that its grid is sufficiently negative with respect to cathode to prevent it striking. The open-ended delay network in the anode circuit of valve 2 charges to $1V_B$ through the large resistance R_4 in series with the comparatively small resistance R_7 . Valve 3 is biased to cut-off by the voltage $-V_G$.

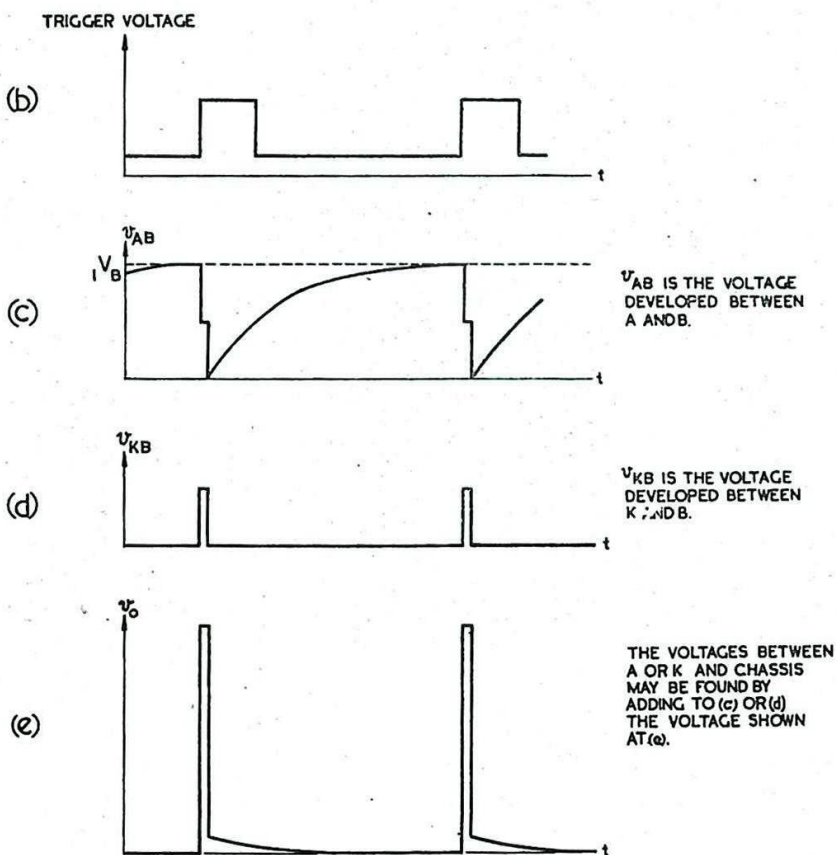
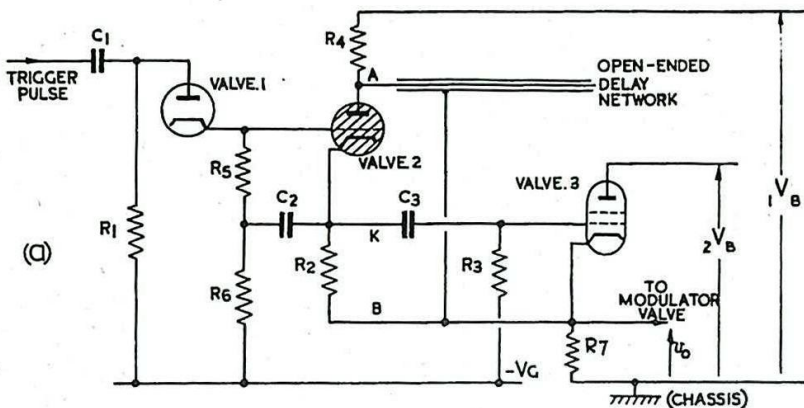


Fig.620 - Boot-strap pre-modulator.

The circuit can be triggered by any pulse with a fairly sharp positive-going edge. This pulse is applied to the short time-constant circuit $C_1 - R_1$. The positive-going portion of the output of this network is applied to the grid of valve 2 via the isolating diode, valve 1. Valve 2 strikes and the delay network discharges through the series combination of valve 2, which has negligible resistance, and R_2 , the value of which is made equal to the characteristic impedance of the delay network.

At the instant of striking, the input voltage to the delay network falls to $\frac{1V_B}{2}$, and, neglecting any loss across valve 2, an equal

voltage appears across R_2 . The voltage across R_2 is applied via the coupling components C_3 , R_3 to the grid of valve 3 which is brought into conduction. The flow of current through R_7 raises the cathode potential of valve 3, but this valve does not act as a cathode follower because the lower end of R_2 is returned to its cathode. As the cathode voltage of valve 3 rises, the voltage of R_2 as a whole rises; this raises the cathode voltage of valve 2, but because valve 2 is a gas triode, once struck it continues to conduct, irrespective of its grid-cathode voltage. The voltage across R_2 , i.e. that applied between grid and cathode of valve 3, remains constant, since the anode-cathode circuit of valve 2, including the delay network, rises with the cathode of valve 3. R_4 is large and has little shunting effect on this "bootstrap" action. Provided the voltage across R_2 is of sufficient magnitude to raise the potential at the grid of valve 3 above that of its cathode, this valve can be made to conduct very heavily, the voltage across R_7 approaching the value $2V_B$.

The time for which valve 2 and valve 3 remain in conduction is determined by the transit time of the delay network. At the instant at which valve 2 is struck and the delay network starts to discharge, a negative-going rectangular wave of amplitude $\frac{1V_B}{2}$ commences to travel

along the network and is reflected without change of phase at the open end. When the wave returns to the input end the line is completely discharged and the voltage across valve 2 and R_2 in series falls to zero. Valve 2 is extinguished, the driving voltage for valve 3 disappears and its current is cut off. The output pulse to the modulator valve therefore lasts for the double transit time of the delay network. The delay network subsequently recharges, valve 2 remaining non-conducting until the next triggering pulse is applied.

A condenser C_2 is connected from the cathode of valve 2 to the junction of two resistors R_5 and R_6 in order to prevent the grid-cathode voltage of valve 2 becoming highly negative when the cathode voltage of valve 2 rises during the pulse due to the "bootstrap" action. The purpose of the isolating diode is to prevent the passage of a high positive-going pulse back to the triggering circuit from the grid of the valve due to the presence of C_2 .

The potentials of the supply circuits are chosen with respect to earth potential so that the output pulse to the modulator rises to a suitable value. Since this is normally not greater than earth potential the common negative supply line (shown connected to chassis) is usually several hundreds of volts negative.

22. Single Valve Pre-Modulator

A pre-modulator circuit using a single valve only is shown in Fig. 621 together with the appropriate waveforms. The circuit consists of a triggered blocking oscillator, modified by the addition of a delay network which determines the duration of the pulse produced. The anode and grid circuits of the valve are coupled together by means of two windings of a three-winding pulse transformer, and the output pulse is taken across the third winding. One end of the grid winding is connected to an open-circuited delay network, the circuit to earth being completed via the resistance R_1 , across which the triggering voltage is developed. The grid is normally biased beyond cut-off through the grid leak R_2 .

The circuit can be triggered by the leading edge of a positive-going pulse of sufficient amplitude. This pulse is applied to a short time-constant circuit $C_1 - R_1$, and the output is applied via the delay network and the grid winding of the transformer to the grid of the valve

bringing the voltage above cut-off. Regenerative action then causes the valve to conduct heavily, a large voltage of polarity as indicated in Fig. 621 appearing across the grid winding. This winding now acts as a source of voltage which charges the delay network via the resistance R_1 and the grid-cathode circuit of the valve, the latter being now of low resistance because the voltage of the grid is positive with respect to that of the cathode. As the delay network charges a negative-going rectangular wave travels towards the open end, where it is reflected without change of phase. On its return to the input end, the grid voltage falls and the valve current is cut off. The output pulse thus lasts for the double transit time of the delay network. Damping

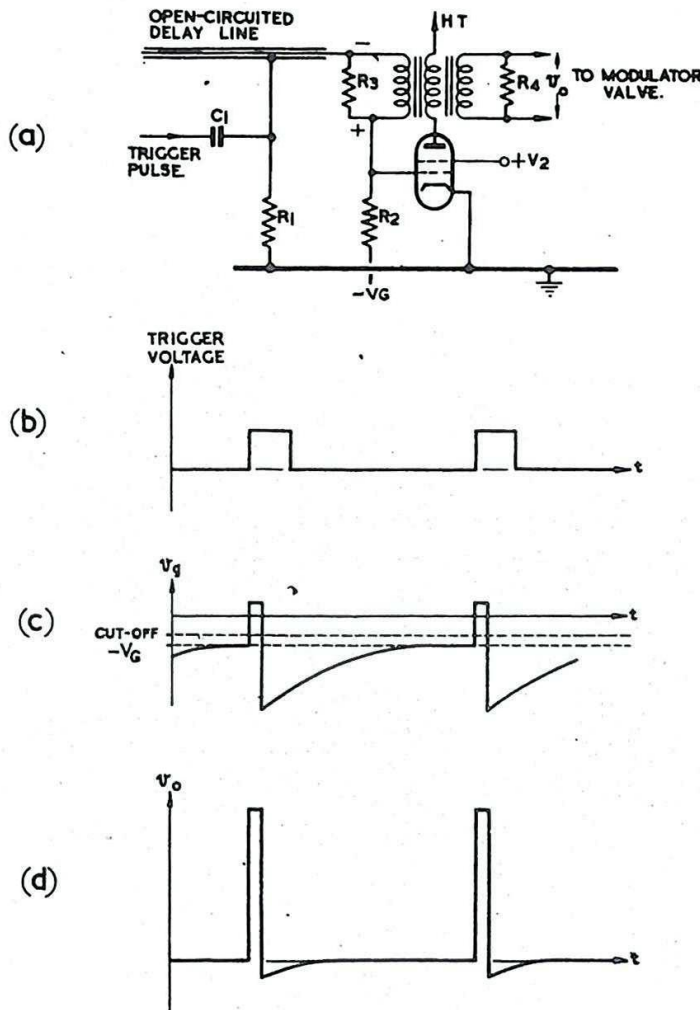


Fig. 621 - Single-valve pre-modulator.

resistors R_3 and R_4 across the windings prevent any appreciable ringing effects. While the valve current is cut off the delay network discharges relatively slowly through the circuit comprising the grid winding (in parallel with R_3), R_2 , the grid bias source and R_1 , in the interval before the next pulse,

The above description of the action of this circuit is an idealised one, complications occurring in practice due to the transformer characteristics.

CHARGING DELAY NETWORK FROM STEADY VOLTAGE SOURCE

23. Choke Charging

If a delay network formed by n L-C sections, is charged from

a steady supply source through a resistor, the maximum network voltage is equal to that of the supply. If a choke L (Fig. 622) is substituted for the resistor, an amplitude double that of the supply may be obtained. The effective charging circuit then consists of a series resonant L - C combination. As shown in Sec. 11, the effective charging capacitance is nC . If the line is initially uncharged, application of the HT shocks the series L - C circuit into an oscillation of period $2\pi\sqrt{nCL}$. Assuming that the series resistance of the choke is negligible, the waveforms of the voltage v_C across the line, the voltage v_L across the choke, and the current i flowing in the circuit are as shown in Fig. 622. The waveforms illustrated in this figure, and in similar succeeding figures, are very much idealised, since in practice considerable ringing usually occurs from the oscillatory circuit formed by the choke (of the order of 100H) and its self-capacitance. This ringing is usually present during each period immediately after the line is discharged. The action is as follows:-

At time $t = 0$ the condensers are considered to be uncharged, i.e. $v_C = 0$. Therefore, since $v_C + v_L = -V_B$,

$$\text{initially} \quad v_L = -V_B;$$

$$\text{but,} \quad v_L = L \frac{di}{dt};$$

$$\therefore \text{at } t = 0, \frac{di}{dt} = \frac{v_L}{L} = \frac{-V_B}{L}.$$

v_L has its maximum negative value at $t = 0$; hence $\frac{di}{dt}$ also has its maximum negative value at this instant. The value of i , however, is zero because L prevents an instantaneous rise of current. Hence, at $t = 0$, $\frac{dv_C}{dt} = 0$.

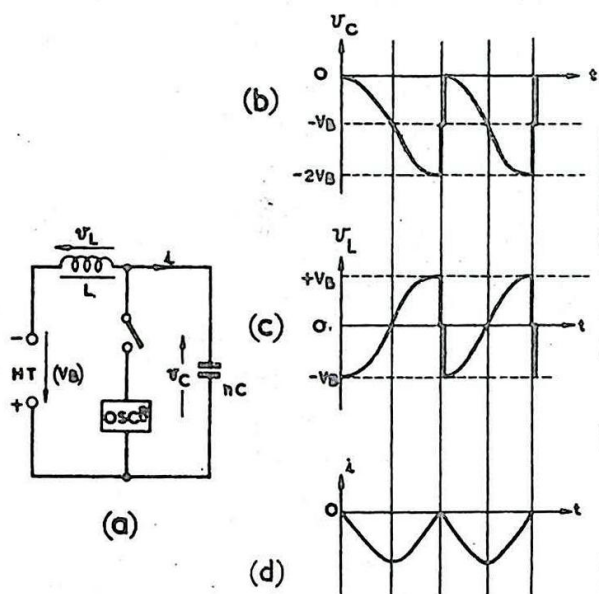


Fig. 622 - Choke charging of delay network.

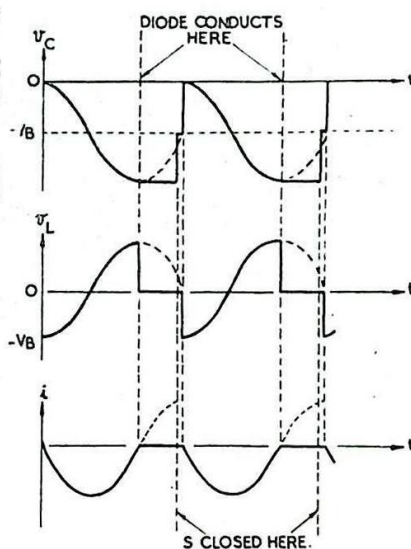


Fig. 623 - Choke charging with hold-off diode.

The circuit oscillates as shown by the waveforms of Fig. 622, until a half-period, $\pi\sqrt{nCL}$ has elapsed.

If at this instant the switch S is closed, the delay network is discharged through the matched load and the line voltage is reduced to zero after the double transit time of the line. Hence if the pulse recurrence period is made equal to $\pi\sqrt{nCL}$, output pulses of amplitude equal to the supply voltage V_B may be obtained.

When this condition is satisfied, we have what is known as Resonant Choke Charging or DC Resonant Charging.

If the recurrence period is greater than $\pi\sqrt{nCL}$ then the line voltage begins to decrease before S is closed. The oscillations during the third quarter cycle are shown dotted in Fig. 623. The reduction in line voltage may be prevented by the insertion between L and the line of a Hold-Off Diode (Fig. 624) which does not allow the current i to reverse, so that v_C is maintained at $-2V_B$ (full lines in Fig. 623).

When the recurrence period is jittery, as for example when a rotary spark gap is employed, its value should be made greater than $\pi\sqrt{nCL}$, and a hold-off diode used. Jitter does not then cause the line voltage to vary.

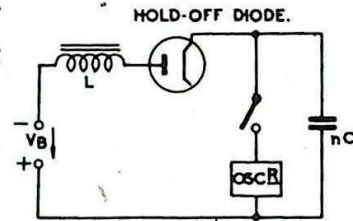


Fig. 624 - Use of hold-off diode.

24. Constant Current Charging

If the recurrence period is less than $\pi\sqrt{nCL}$ then the circuit takes several periods to reach the steady state.

Interval 1

The condensers do not charge fully to $-2V_B$ during the first period (Fig. 625). The current i is still negative when S is closed, not yet having fallen to zero.

Interval 2.

At the start, i has approximately the same value as at the end of interval 1. Therefore, the charging rate $\frac{dv_C}{dt}$ at the start of interval 2 is the same as at the end of the first interval, whereas at the start of that interval it was zero. $\frac{di}{dt}$ is, however, the same

at the beginning of intervals 1 and 2, because in each case $v_C = 0$ and $v_L = L \frac{di}{dt} = -V_B$. Consequently,

not only does the charging start at a higher rate than in interval 1, but, at least initially, the charging rate increases with time. The result of this is that v_C attains a much larger (negative) value by the end of interval 2. This (negative) value may be

(i) equal to,

(ii) less than

or (iii) greater than $-2V_B$,

depending on the relative values of the recurrence period and of

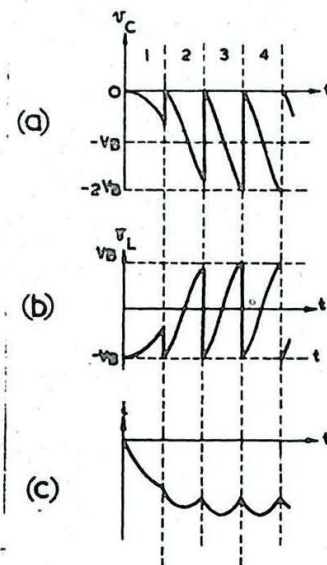


Fig. 625 - Constant current charging of delay network (discharge assumed instantaneous).

$\pi\sqrt{nCL}$. Fig. 625 shows the waveforms for case (ii).

Interval 3

(i) If v_C reaches $-2V_B$ by the end of interval 2, then the current flowing is the same at the start and finish of interval 2, the conditions at the start of intervals 2 and 3 are the same, and the action is repeated.

(ii) If v_C does not reach $-2V_B$ in interval 2, the current flowing at the end of interval 2 is greater than at the start of that interval. Therefore, the current and charging rate at the start of interval 3 are greater than at the start of the second interval and the (negative) value of v_C reached by the end of the interval will be greater.

(iii) If the (negative) value of v_C is greater than $-2V_B$, the current flowing at the end of interval 2 is less than at the start of that interval. Consequently the (negative) value of v_C reached by the end of interval 3 will be less than at the end of the second interval.

Succeeding Intervals

The process is repeated, a steady state being reached when v_C attains the value of $-2V_B$ at the end of each interval. When this occurs the values of i at the beginning and end of each interval become equal, and the current variation is a relatively small one about a constant level. It is from this fact that the term Constant Current Charging is derived.

The actual value of the pulse recurrence period does not affect the steady state condition that $v_C = -2V_B$ at the end of each interval provided that its value is steady. Constant current charging is not suitable for use with a jittery recurrence period.

25. Symmetrical Charging

It has been shown that by charging a delay network through a choke a pulse of amplitude approximately equal to V_B may be obtained. However, with the method described above, it is necessary for the delay network to be capable of withstanding $2V_B$, since it becomes charged to this voltage. This can be avoided by arranging that the voltage across the delay network swings equal amounts above and below zero. Fig. 626.(a) shows a circuit whereby this action may be produced. In this circuit the network charges rapidly during the intervals when the switch S is closed and discharges slowly through L while S is open.

Suppose that the condensers of the delay network are initially

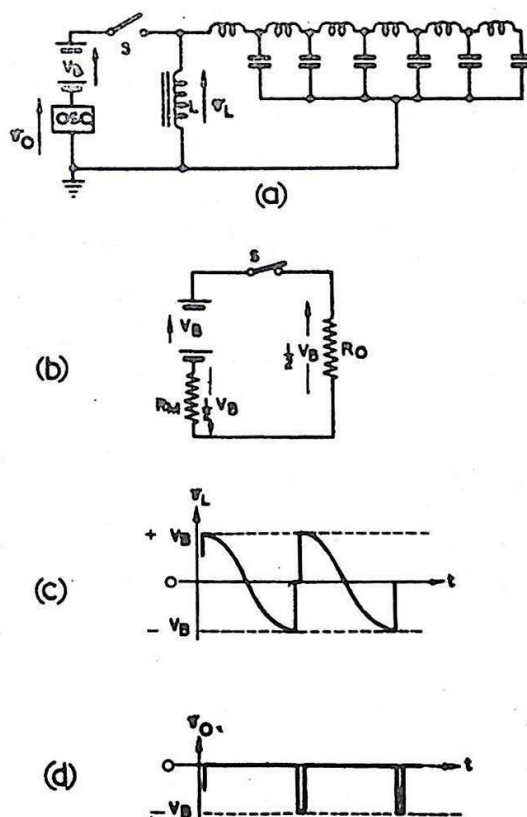


Fig. 626 - Symmetrical charging of delay network.

uncharged. When S is closed V_B is applied to the network, of characteristic impedance R_0 , and the oscillator, of resistance R_M , in series (Fig. 626(b)). If $R_0 = R_M$, a voltage $\frac{1}{2}V_B$ appears across R_0 , and an equal negative voltage across R_M . As the network charges a rectangular wave of amplitude $\frac{1}{2}V_B$ travels along the line.

The wave is reflected without change of phase at the open end and returns to the input end, the voltage across the line increasing to V_B . If $R_0 = R_M$ the wave is completely absorbed at the input end, the charging current falls to zero, and S opens.

The line now discharges through L (Fig. 627(a)). Considering first that the recurrence period is equal to $\pi\sqrt{nCL}$, the voltage across the network swings from $+V_B$ to $-V_B$ during the discharge as shown in Fig. 626(c). The second time that S is closed, the effective voltage acting in the circuit is equal to $+2V_B$ (Fig. 627(b)), and if $R_0 = R_M$ the voltage appearing across the oscillator is equal to $+V_B$. As the delay network charges a rectangular wave of amplitude $+V_B$ travels along the network, the voltage across the line changing from $-V_B$ to zero on the forward journey and from zero to $+V_B$ on the return. The action is then repeated.

If the recurrence period is less than $\pi\sqrt{nCL}$ a constant current form of discharge builds up in the way described in Sec. 24, and the voltage across the line again swings from $+V_B$ to $-V_B$ as for the resonant case.

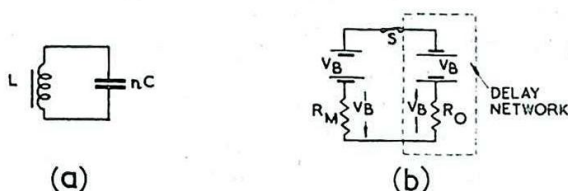


Fig. 627 - Equivalent charging (a) and discharging (b) circuits.

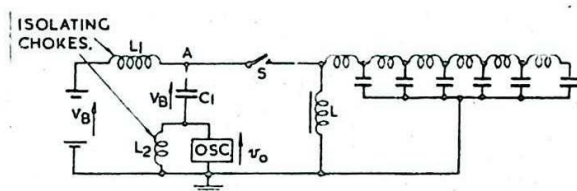


Fig. 628 - Symmetrical charging of delay network: practical circuit.

This circuit therefore generates pulses of amplitude equal to V_B without the voltage across the delay network exceeding this value, but in its present form neither side of the HT supply can be earthed. This disadvantage may be overcome by re-arranging the circuit as shown in Fig. 628. The oscillator, shunted by a coil L_2 , is inserted in series with the smoothing components L_1 , C_1 of the HT supply. The circuit comprising the HT supply, L_1 and C_1 , is completed by L_2 during the intervals when the oscillator is not conducting. The Isolating Inductances, L_1 and L_2 , are both small compared with L , but are of sufficient inductance to act during the charging process as virtual open-circuits. The action of the circuit is similar in every way to that described immediately above, the condenser C_1 taking the place of the HT supply, the time-constant $C_1 (R_M + R_0)$ being very much greater than the pulse duration.

26. TYPICAL SPARK GAP MODULATOR

The circuit of a typical spark gap modulator is shown in Fig. 629.

Valves 1 and 2 form a multivibrator, producing a 20-microsecond pulse output, which is fed to the grid of the trigger valve, valve 3. This valve is normally biased beyond cut-off, but is brought into heavy conduction by the positive-going 20 microsecond pulse. On the trailing edge of this pulse, valve 3 is cut-off, and

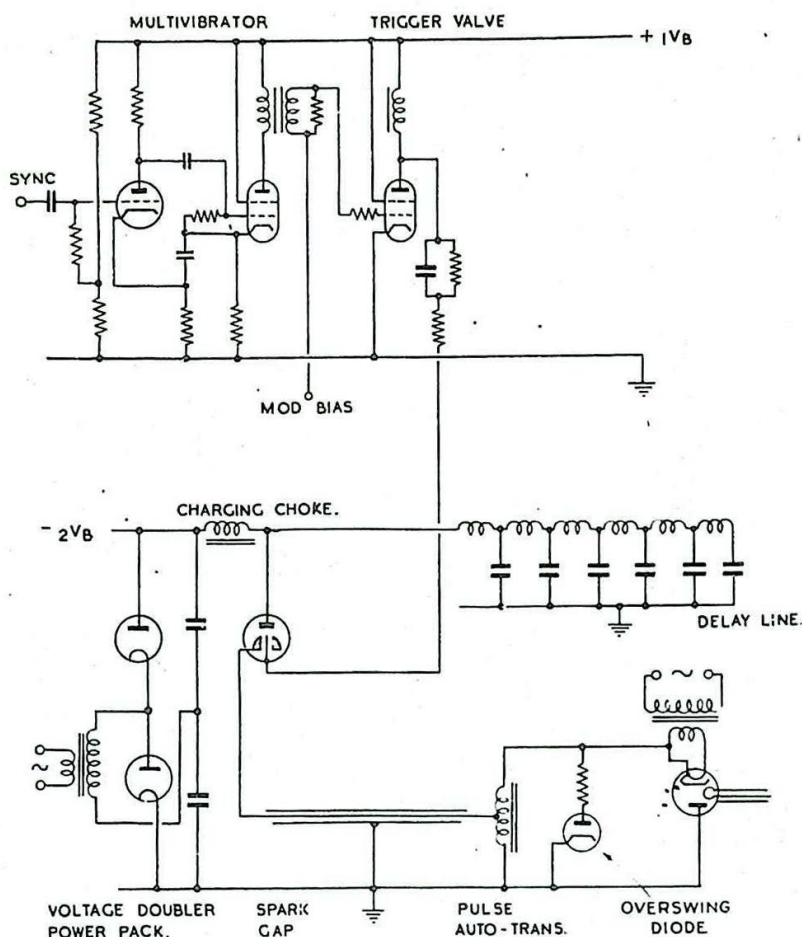


Fig. 629 - Typical spark-gap modulator.

its anode circuit rings, the first positive-going swing being of large amplitude. This is used to trigger the spark gap, which discharges the delay network. The 1-microsecond pulse so formed is fed to the magnetron via a matched connecting cable, and a step-up pulse transformer. The latter matches the cable to the load presented by the magnetron when conducting, and converts the pulse from 4kV, 40A to 16kV, 10A. The delay network is charged through a choke from a voltage doubler power pack.

CHARGING DELAY NETWORK FROM AN ALTERNATING VOLTAGE SOURCE

27. Advantages of Alternator Charging

Although it is simpler to charge a line modulator from a source of steady voltage, there are definite advantages to be gained from the use of an unrectified alternating supply. The principal advantages are:-

- (i) No high voltage rectifier is required.
- (ii) No high voltage smoothing circuit is needed.
- (iii) The high voltage transformer need not be so large.
- (iv) A hold-off diode is unnecessary (see Sec. 23).
- (v) The maximum HT voltage is developed across the line for only a small fraction of the recurrence period. This helps to ensure that the transmitter does not oscillate except when required.

These advantages considerably reduce the bulk and weight of the modulator unit. However, this method of charging is usable only if the alternator frequency is constant within narrow limits. This frequency is also the recurrence frequency of the output pulses.

28. The Charging Circuit

Fig. 630(a) shows a simplified diagram of a circuit for charging a line modulator from an alternator. The line discharging circuit (b) is connected between A and B. While the discharge valve is not conducting AB is open-circuited, and, to the alternating voltage of the frequency normally used, of the order 500-1000 c/s, the impedance of the line to the right of AB is a small capacitance C. The input impedance of the line transformer T_1 may then be shown to be approximately $\frac{-j}{m^2 \omega C}$ where m is the step-up ratio of the transformer and $\omega = 2 \pi f_a$, f_a being the alternator frequency. This result neglects the magnetising inductance of the transformer. A more accurate result gives the input admittance of T_1 as the sum of the susceptances of the magnetising inductance and of the input capacitance $m^2 C$.

*** Input Impedance of Transformer when Secondary Load is a small Capacitance

The input impedance z_t of the transformer T_1 terminated in C is given by

$$z_t = j\omega L_1 + \frac{\omega^2 M^2}{j\omega L_2 + \frac{1}{j\omega C}}, \quad (\text{see Chap. 1 Sec. 20})$$

where L_1 , L_2 are primary and secondary inductances respectively and M is the mutual inductance between primary and secondary windings.

Ideally, $\omega L_2 \gg \frac{1}{\omega C}$, $M^2 \doteq L_1 L_2$, and $\frac{L_2}{L_1} \doteq m^2$ where m is

the transformer step-up ratio.

The equation for z_t simplifies as follows :-

$$\begin{aligned} z_t &\doteq j\omega L_1 - \frac{j\omega^2 L_1 L_2}{\omega L_2 - \frac{1}{\omega C}} \\ &\doteq j\omega L_1 - \frac{j\omega^2 L_1 L_2}{\omega L_2 (1 - \frac{1}{\omega^2 C L_2})} \\ &\doteq j\omega L_1 - j\omega L_1 (1 + \frac{1}{\omega^2 C L_2}) \\ &= - \frac{j\omega L_1}{\omega^2 C L_2} \\ &\doteq - \frac{1}{\omega C} \cdot \frac{1}{m^2} \end{aligned}$$

Hence the transformer, when terminated in C , "looks like" a capacitance m^2C .

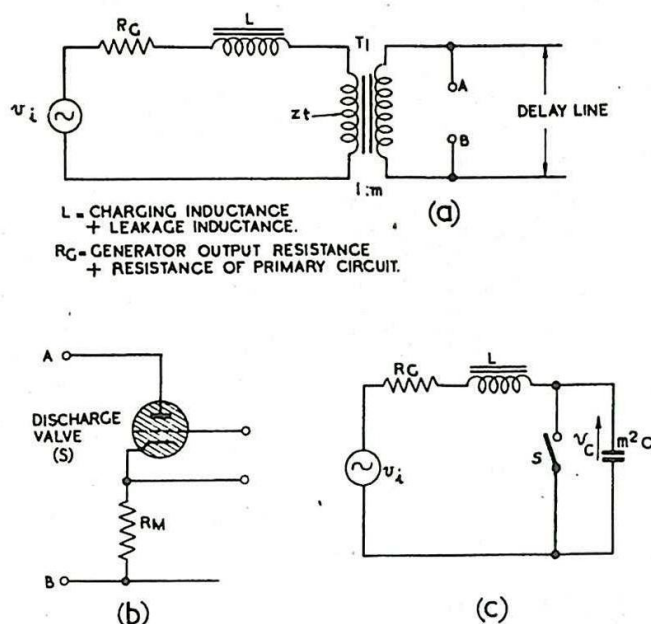


Fig. 630 - Alternator charging: basic circuits.

29. The Equivalent Charging Circuit

Neglecting the magnetising inductance of the transformer we may represent the charging circuit by the equivalent arrangement shown in Fig. 630(c).

When the switch S is closed the line discharges through the oscillator load R_M . This discharge action is described in Chap. 4, Sec. 12. For a resistive load the line may be assumed to be completely discharged in a few microseconds, even if it is not properly matched to the load. So far as the low frequency charging circuit is concerned, this discharge may be considered as instantaneous, and may be represented by the momentary short-circuiting of the terminals AB (Fig. 630(a)).

In the case of a magnetron load, and in some other instances, an overswing generally occurs. The reasons for this overswing, and its effect, are discussed later; for the present it is assumed that the line is discharged completely during the momentary closure of the switch S (Fig. 630(c)) representing the firing of the discharge valve.

30. Transients Due to Line Discharge

The closure of the switch S (Fig. 630(c)) reduces the impedance across the effective line capacitance m^2C to zero (the justification for this may be found by putting zero in place of $\frac{1}{\omega C}$ in the

analysis given in Sec. 28). The corresponding input voltage v_c also becomes zero. Since the short-circuiting is only of momentary duration no appreciable change in current occurs in the primary circuit during the discharge of the line, the time constant $\frac{L}{R_G}$ being much greater than the duration of the discharge.

An interval follows during which the circuit passes through a transient state and, in the absence of further irregularities, tends

gradually towards the steady state, or normal AC working condition.

In the actual circuit this steady state is never reached, but for ease of explanation it will be assumed in what follows, in the first instance, that the switch is closed once only, when the line voltage is at any portion of its steady state condition (Fig. 631). Later it will be described how the line voltage is maintained in a condition little resembling a steady sinusoidal form.

31. Graphical Derivation of Transient State

The manner in which the line voltage returns to its steady state after a single momentary discharge is illustrated in Fig. 631(a). This picture is derived as follows.

The steady state voltage across m^2C (Fig. 630(c)) is indicated by the full curve from O to E and from then on by the dotted curve. It is to this steady state that the transient conditions ultimately return. The actual voltage v_C is shown by the full line.

When the line is momentarily discharged (E) v_C falls to zero

(F). The line current given by $i = m^2C \frac{dv_C}{dt}$

is unchanged, so that $\frac{dv_C}{dt}$ has the same value immediately after the

discharge as it had immediately before. The gradients of the pre-discharge and post-discharge curves at E and F respectively are therefore the same.

The resultant line voltage after the discharge may be obtained by adding to the steady state solution (dotted curve Fig. 631(a)) the appropriate transient (Fig. 631(b)). (This corresponds to the addition of the complementary function (transient solution) to the particular integral (steady state solution) in the solution of the linear differential equation of the system).

To obtain this appropriate transient, first draw the characteristic curve of the response of the series ringing circuit, formed by R_C , L and m^2C , to a sudden change in input voltage. This is illustrated in Fig. 632(a). This curve is one of an infinite sequence of curves, some of which are shown in Fig. 632(b). From such a set of curves choose the one which satisfies the conditions, i.e. having the right magnitude and slope.

In the case illustrated in Fig. 631 the initial magnitude is equal to EF, its value is negative, and the initial slope is zero. It is this curve, starting from the point H, where it meets PQ (Figs. 632(b) and 631(b)) which is the correct transient for the case considered. When this is added to the steady state solution (Fig. 631(a)) the resultant amplitude (zero) and gradient (shown at E and F) are in accordance with the initial conditions.

The resultant curve is thus obtained (full curve, Fig. 631(a)).

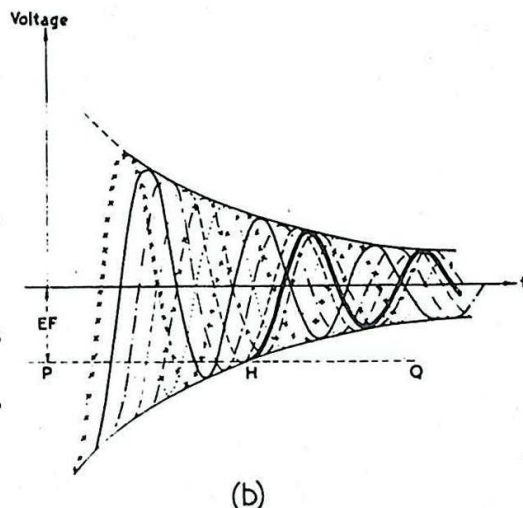
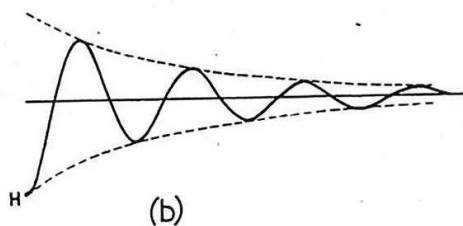
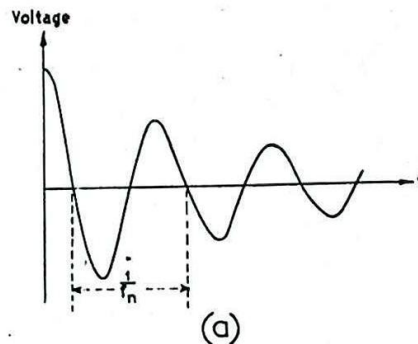
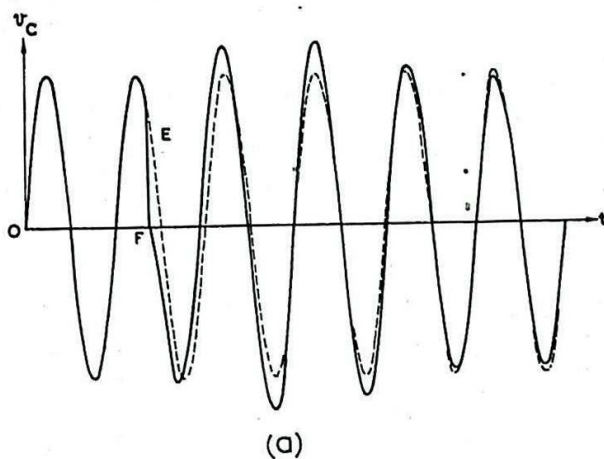


Fig. 631 - Transient effect of short-circuiting the transformer T_1 .

Fig. 632 - Determination of the correct transient.

32. Normal Transient Conditions

In practice the circuit is not allowed to return to the steady state. The line discharge valve is triggered each time the line voltage reaches a maximum and the voltage v_c varies as shown in Fig. 633. Since the voltage is a maximum the current $i = m^2 C \frac{dv_c}{dt}$ is zero,

both immediately before and immediately after the discharge. The transient curve which must be added to the steady state curve to give v_c must therefore satisfy the following conditions:-

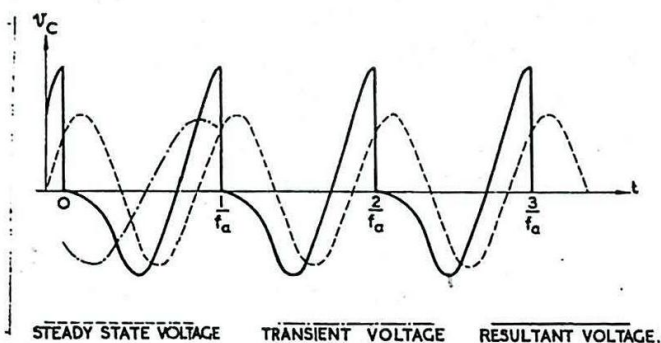


Fig. 633 - Line charging curve.

- (i) It must have the same magnitude as the steady state voltage at the moment of striking ($t = 0$) and be of opposite polarity.
- (ii) It must have a gradient equal and opposite to that of the steady state curve at $t = 0$.
- (iii) When it is added to the steady state voltage the resultant must have a maximum value at $t = \frac{1}{f_a}$, where f_a is the alternator frequency.

After $t = \frac{1}{f_a}$ the line discharges again and the cycle repeats (Fig. 633).

The form of the line voltage depends chiefly on the ratio f_n/f_a , where f_n is the natural frequency of the series resonant circuit formed by R_G , L and m^2C . The optimum value for this ratio is dealt with in Sec. 34. The value $\frac{f_n}{f_a} = 0.7$ is usually chosen,

and this is the value used for the charging curve of Fig. 633.

33. Effect of Overswing

If for any reason an overswing occurs (see Sec. 35) the "initial" voltage v_c immediately following each discharge of the line is not zero, but is $-kv_c$, where \hat{v}_c is the value of v_c immediately before the discharge and k is the overswing ratio. This causes the form of v_c to be changed from that of Fig.

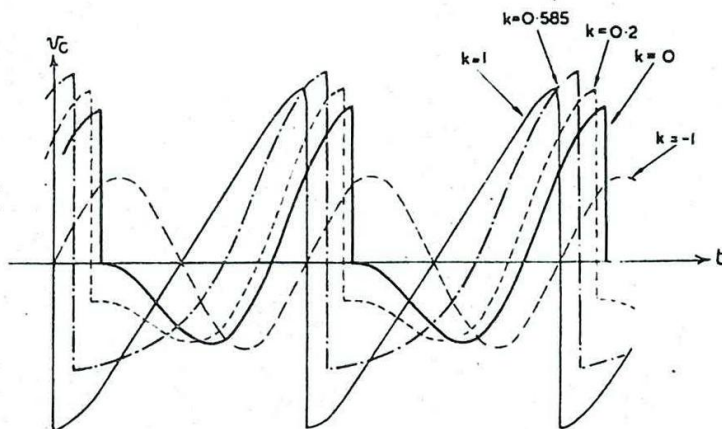


Fig. 634 - Effect of overswing, showing good regulation ($f_n/f_a = 0.7$).

633, as indicated in Fig. 634. The instant at which the line discharges, relative to the alternator voltage (or to the steady state voltage at the line transformer) changes so that the new conditions are maintained for each discharge of the line. The phase angle at which the line discharges relative to the alternator, is called the Angle of Fire. A change in the overswing ratio k thus causes a change in the angle of fire, as shown in Fig. 634. A small overswing is advantageous in the case of most transmitters, as this holds the transmitter valve supply voltage at the opposite polarity to that required for oscillation, and facilitates a rapid dying away of the output pulse. It is possible, however, for overswing to cause the line voltage to acquire dangerously large values.

34. Choice of the Ratio f_n/f_a

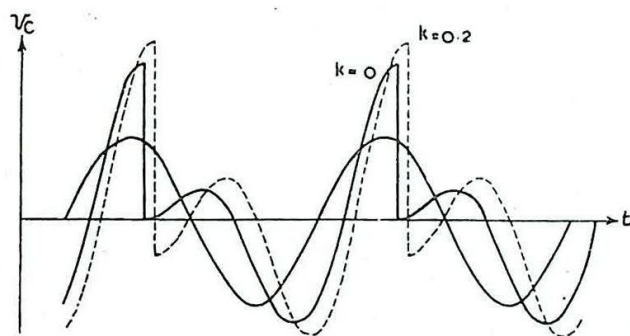
The effect of the ratio f_n/f_a upon the charging curves is complicated and may be divided into three parts:-

- (i) It determines the angle of fire and the shape of the charging curve, i.e. the form of the line voltage.
- (ii) It determines the ratio of the line voltage to the alternator EMF, since this depends on the frequency of operation f_a in comparison with the resonant frequency $\frac{1}{2\pi\sqrt{LC}}$, and this is very nearly the same as f_n . In general, operation near resonance means that only a small alternator EMF is required.
- (iii) It affects the reaction of the circuit to overswings. These may occur through the normal operation, in which case they are expected and can be allowed for, or they may arise through abnormal loading conditions, such as a flash-over, or an open-circuit due to faulty connection. The effect on the line voltage of an overswing determines the Regulation of the circuit. If the line voltage increases considerably when an overswing occurs the regulation is said to be poor.

The various values for f_n/f_a between 0 and 2 are divided into five sections as under:-

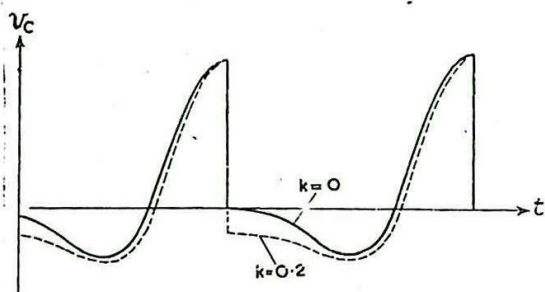
f_n/f_a	Remarks
(a) 0 - 0.5	Requires large alternator EMF
(b) 0.5 - 0.8	Optimum conditions
(c) 0.8 - 1.3	Poor regulation
(d) 1.3 - 1.6	Good regulation but charging curve less favourable than that of (b) (see Fig. 635(a)).
(e) 1.6 - 2.0	Ill-shaped charging curve, poor regulation, large alternator EMF.

Fig. 635 illustrates some of these points.



(a) $f_n/f_a = 1.4$ $k=0$ (RESISTIVE LOAD)
 $k=0.2$ (MAGNETRON LOAD)

Fig. 635 - Line charging curves.



(b) $f_n/f_a = 0.44$ $k=0$ (RESISTIVE LOAD)
 $k=0.2$ (MAGNETRON LOAD)

35. Causes of Overswing

When the output from a line-modulator is used to pulse a magnetron it is unlikely that the whole of the energy of the modulating pulse is absorbed by the transmitter. This is particularly so if the pulse shape departs markedly from the ideal rectangular form, causing the magnetron impedance at the beginning and end of the pulse to change so that the line is mismatched. This may cause the line to reverse its polarity (see Chap. 4, Sec. 12), subsequent reversals being prevented by the cessation of ionisation in the discharge valve (assumed soft).

If the load becomes short-circuited, the tendency to over-swing depends on whether spark gaps (unpolarised discharger valves) or gas-filled triodes (polarised discharger valves) are used. In the former case the line energy would probably be rapidly absorbed by the spark gap and other circuit elements during the oscillatory discharge, and dissipated as heat. In the latter case the valve would probably open-circuit following a complete overswing ($k = 1$) so that the charging curve would be considerably modified.

If the line is open-circuited so that the discharger valve does not operate the transient state never develops ($k = 1$). If, however, a pulse transformer is used between the line and the transmitter an open-circuit at the transmitter means that the cable discharges through the magnetising inductance of the pulse transformer, with considerable dissipation of energy in the transformer core.

36. Method of Triggering Discharge Valve

Fig. 636 shows a method of triggering the discharge valve at each instant when the line voltage is a maximum. The current i through the primary winding of the transformer T_1 , represented by m^2C in Fig. 630(b), is given by the relation

$$i = m^2C \frac{dv_C}{dt}.$$

Hence when v_C is a maximum i is zero. This current is caused to flow through the primary winding of a peaking transformer T_2 (Fig. 636(a)). This is a transformer with a small iron core which saturates for small values of primary current. The variation of flux corresponding to the primary current (Fig. 636(b)) is shown at (c) and the secondary EMF is indicated at (d). (EMF \propto rate of change of primary flux). The output voltage v_p developed across the secondary winding of this transformer is thus of a form suitable for triggering the discharge valve at the instant the line voltage reaches a maximum.

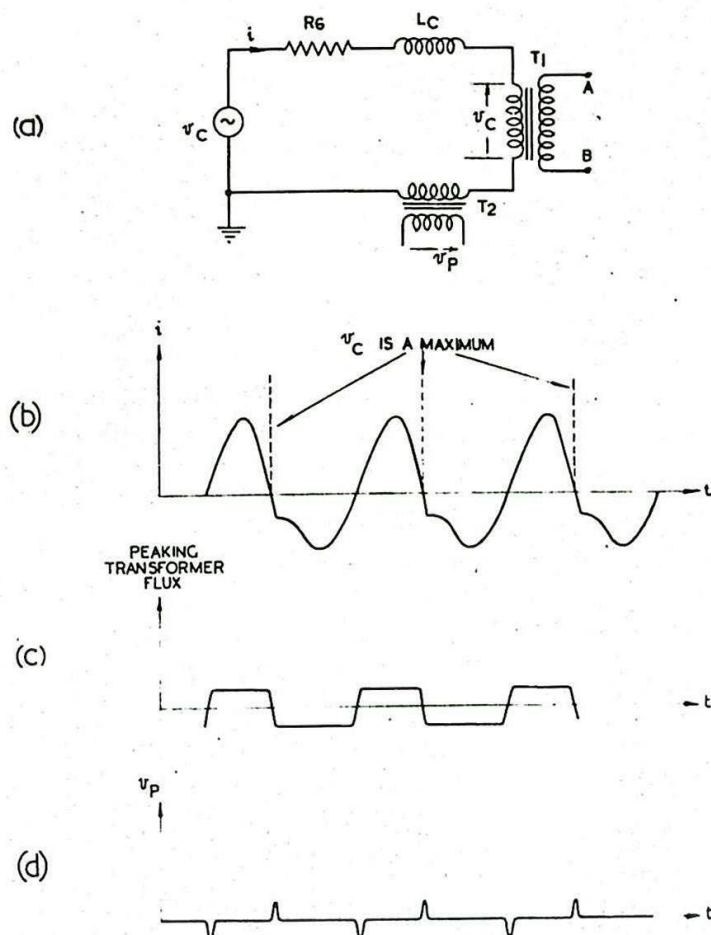


Fig. 636 - Derivation of triggering pulses for discharge valve.

VOLTAGE MULTIPLYING CIRCUITS

37. General

It is often necessary to avoid the use of high voltage supply circuits by employing voltage multiplying devices. One such device, the pulse transformer, has already been described. Other methods, involving the use of more than one pulse-forming network, are described below. They operate on the principle of being charged in parallel from the same supply and, by means of synchronous switches, discharged in series with the load. The Marx circuit is described in schematic form only. The Blumlein circuit, which is a derivative of the Marx, is dealt with more fully because its application in radar transmitting circuits is already established.

38. The Marx Circuit

The Marx Circuit is illustrated in Fig. 637.

C_1, C_2, \dots, C_n represent the charging capacitances of n separate networks. When the switches $S_1, S_2, S_3, \dots, S_n$ are open, $C_1, C_2, C_3, \dots, C_n$ are charged in parallel from the supply (assumed, for simplicity, to be a steady voltage source) through the charging choke L and the isolating chokes L_1, L_2, L_3 etc. The inductances L_1, L_2, L_3 etc. are all small compared with L and to a first

approximation their effect on the charging process may be neglected.

If all the switches S , S_1 , S_2 etc. are closed simultaneously, with the networks thus charged, the latter are instantaneously connected in series and commence to discharge through the load, R_M . Provided each inductance is $\gg C_T R_M^2$, where C_T is the total effective capacitance of C_1, C_2, \dots etc. in series, it has little effect on the discharge, which takes the path of short time-constant $C_T R_M$ in preference to that of the long time-constant $\frac{L_T}{R_M}$, (L_T is the total

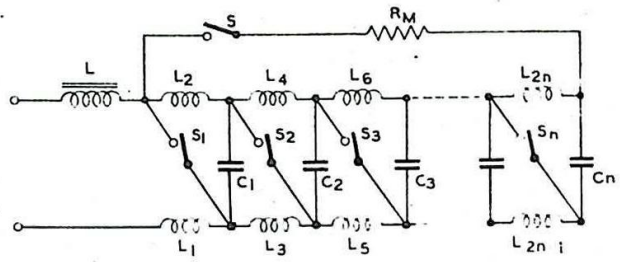


Fig. 637. - Marx circuit.

effective inductance of L_1, L_2 etc. when the switches are closed. Thus the lines are charged in parallel and discharge in series with the load. The discharge is similar to that of a single line of characteristic impedance equal to the sum of the characteristic impedance of the individual lines i.e. $n R_0$, and charged to a voltage $n \times$ (the voltage on a single line). The inductances L_1, L_2 etc. are usually made equal and intermediate in value between the total series inductance of each line and the inductance L of the choke.

The switches S, S_1 , etc. may take the form of triggered spark gaps or gas-filled triodes. The principal disadvantage of this method arises from the difficulty of synchronising the switching.

39. Blumlein Modulator

The derivative of the Marx circuit most commonly employed in Service radar is the Blumlein modulator which employs two delay lines.

To avoid the difficulty of triggering simultaneously the two switches necessitated by the Marx circuit, this variant was designed to operate with a single spark gap.

The basic Blumlein circuit is illustrated in Fig. 638. The lines are charged in parallel with the switch S open. During the discharge the generator is isolated from the network by the choke L and will be ignored in the subsequent analysis.

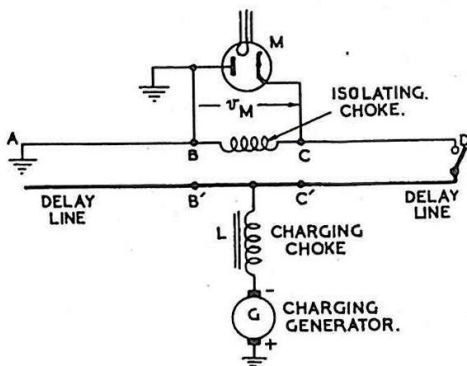


Fig. 638 - Blumlein circuit.

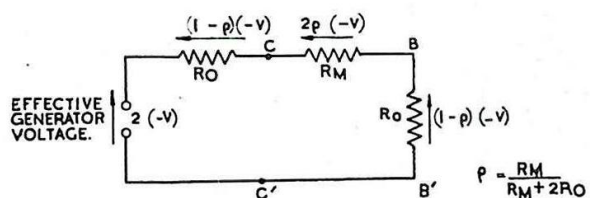


Fig. 639 - Equivalent circuit showing voltage developed across the load by a wave $(-V)$ arriving at the termination CC' .

Suppose that the two lines are of identical construction, lossless and with characteristic impedance R_0 , and transit time T . The magnetron load M has resistance R_M to a pulse of the right voltage and polarity to ensure that it oscillates correctly. Let the lines be charged to a voltage V when the switch S is closed at $t = 0$. This closure short-circuits the line CD so that a wave of magnitude $-V$ travels towards C . This divides at C into three parts:-

- (i) a fraction ρ is reflected towards D .
- (ii) a fraction 2ρ is developed across the load.
- (iii) a fraction $1-\rho$ is developed across the input to the line BA .

Thus immediately after the arrival of the wave at C , the voltages developed across the respective terminals due to the wave are as shown in Fig. 639. This result may be derived by the method indicated in Chap. 4, Sec. 9. The value of ρ is given by

$$\rho = \frac{R_M + R_0 - R_0}{R_M + R_0 + R_0} = \frac{R_M}{R_M + 2R_0}.$$

The sequence of waves and load voltages is illustrated in Fig. 640. It is assumed that the switch remains short-circuited throughout the action.

For a wave travelling from A to B , the division at BB' is as follows:-

- (iv) a fraction ρ is reflected towards A .
- (v) a fraction -2ρ is delivered to the load.
- (vi) a fraction $1-\rho$ is developed across CC' .

[The reversal of sign of the voltage delivered to the load, compared with the former case of the wave travelling from D to C , is due to the fact that it is necessary to have a uniform sign convention for the voltage across the load. If this is positive when approached from D it is negative when approached from A .] Thus a pulse of magnitude $2\rho V$ is developed across the magnetron at the instant $t = T$. This pulse must be maintained at this magnitude until $t = 3T$, by which time the waves travelling along the lines CD and BA will have undergone reflections and the

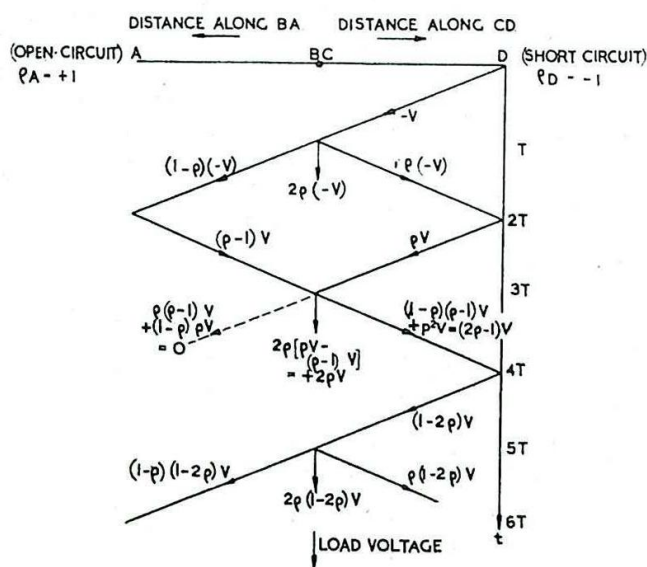


Fig. 640 - Action of Blumlein modulator.

wave fronts will have returned to the load.

The figure shows the passage of various wave fronts along the lines, suffering simultaneous reflections at the ends A and D, and returning to B and C, where fresh pulses are developed across the load. The load voltages are indicated by the vertical lines. The amplitude and polarity of the various waves are indicated by the symbols written on the appropriate lines.

Reflections at the short-circuit D involve a sign-reversal; at the open-circuit A there is no change of polarity of the reflected wave.

Further, the values shown in the diagram are incremental values. Thus at $t = T$ a pulse $-2\rho V$ is developed across the transmitter. The increment $+2\rho V$ at $t = 3T$ raises the load voltage to $-2\rho V + 2\rho V = 0$ at which value it remains for a further interval $2T$, and so on.

The load voltage is illustrated in Fig. 641. This ideal form is not likely to be achieved in

practice because of changes of the magnetron impedance due to the non-rectangularity of the applied pulse, unwanted reflections, etc.

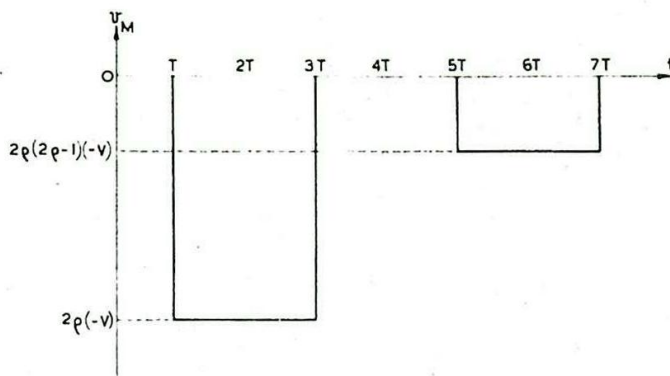
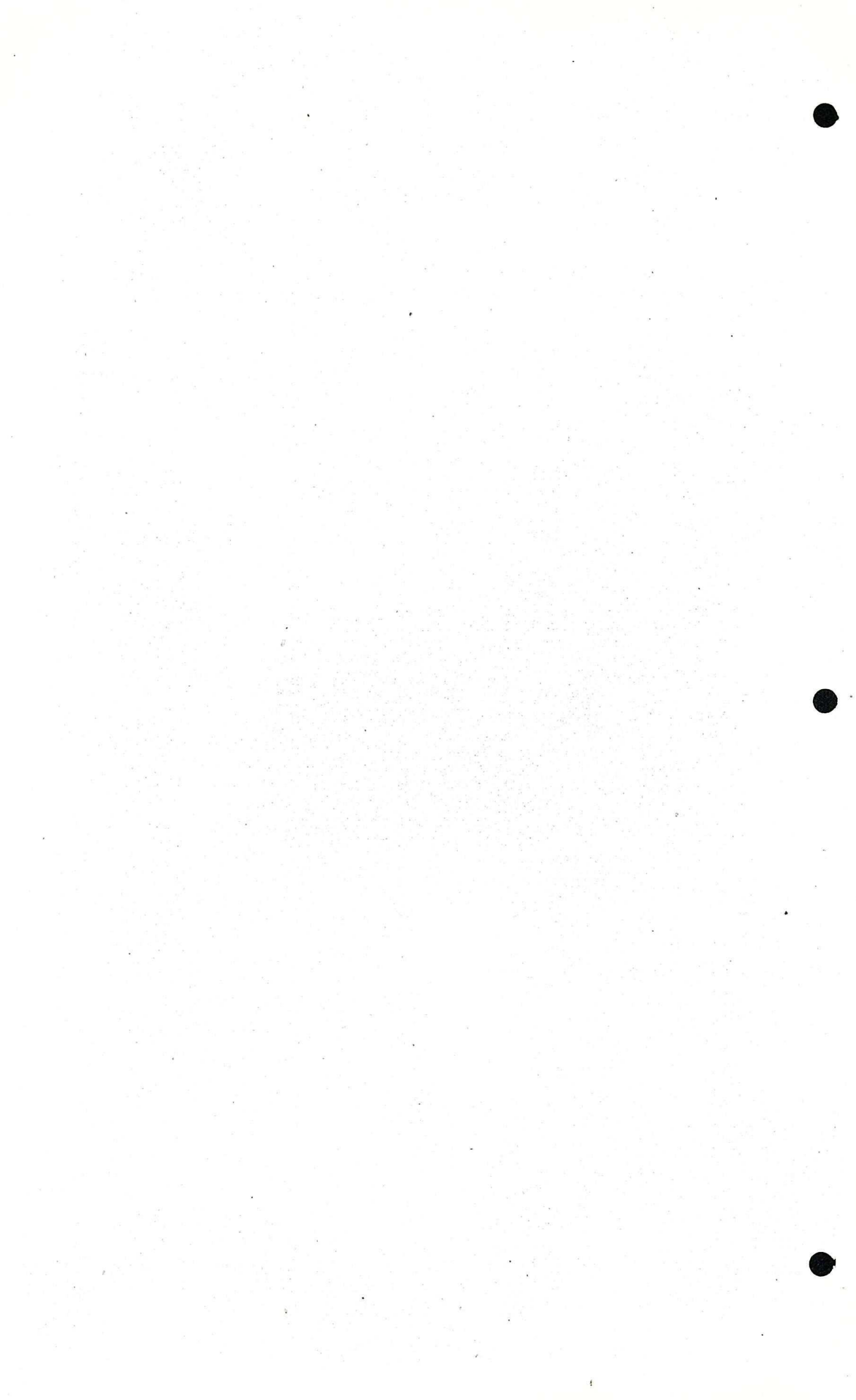


Fig. 641 - Load voltage: (the second pulse is positive if $\rho < 0.5$; i.e. if $R_M < 2R_0$.)

In the Blumlein circuit the magnetron is matched to the network when $R_M = 2R_0$. In this case, provided the pulse were perfectly rectangular and the magnetron impedance when oscillating were constant and equal to R_M the lines would be completely discharged after time $t = 3T$ from the instant S is closed.



CHAPTER 15

NOISE

1. INTRODUCTION

The general term Noise covers all undesirable electrical disturbances, either externally or internally generated, which tend to mask the clear reception of signals. In communications noise is often referred to as "mush", and in radar, by virtue of the appearance of the trace it produces on the screen of a CRT, as "grass".

Noise may first be divided into two general classes:-

(i) Avoidable noise

Examples of noise sources within this category are:-

- (a) "Man-made" static arising from generators, ignition systems, etc.
- (b) Faulty contacts in receivers.
- (c) Poor quality resistors in receivers.

Avoidable noise is not considered further, since it is obvious that no theory or general prediction can be made in regard to such noise. These effects can be minimised by improved manufacture, use of high-quality components, systematic servicing, satisfactory weather protection, and care in the siting of equipments.

(ii) Unavoidable noise

This noise may be divided into two sub-classes:-

(a) External noise

Atmospheric noise, chiefly due to thunderstorms, signals from whose discharges may be propagated over very great distances. Radiation of cosmic origin.

(b) Internal noise

This class of noise covers all noise generated within the receiver, after exclusion of all avoidable internal noise due to bad contacts, etc; and in addition includes thermal noise generated in the aerial.

The subject matter of this chapter is mainly concerned with internal noise.

2. SIGNAL-TO-NOISE RATIO

Noise, from whatever source, after passing through the detector stage of a receiver presents itself on the time-base of a CRT as a ragged but very characteristic signal deflection, extending across the screen with more or less uniform amplitude (Fig. 642). On many types of display (e.g. meter), noise appears as a random variation of the position of the indicating device (pointer, electron beam, etc.) whose instantaneous position at any given moment of time may vary between wide limits, but which has a more or less clearly defined mean amplitude over any appreciable time interval.

In any element of a receiver the noise in the absence of signals has a certain mean power, and the signal in the absence of noise presents an average power for its duration. Calling these powers Noise Power and Signal Power respectively, then the ratio of signal power to noise power is called the Signal-to-Noise Ratio at that point in the receiver.

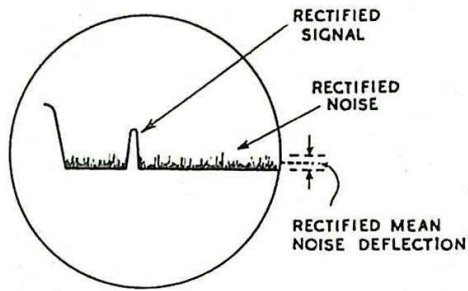


Fig. 642 - Signal and noise at CRT (A-type display).

Fig. 642 shows a typical A-type Display. As a useful guide it may be noted that a pulse signal is just distinguished on such a display by its amplitude rising above the noise level when the signal-to-noise ratio just prior to the detector of the receiver is about unity. If the ratio could be examined at previous points in the receiver (and in particular at the input to the first stage), it would be found that the ratio would increase as the input stage was approached. In the case of a centimetre equipment a ratio of signal-to-noise of between 25:1 and 100:1 must be present at the input for unity ratio just prior to the detector. In other words, to achieve a just distinguishable indication of signal on an average A-type display it is necessary to inject a signal power of 25 to 100 times the noise power existing at the input to the first stage.

3. INTERNAL NOISE

The sources of internal noise at present known to occur in normal circuits may be broadly divided into the following classes:-

- (i) Thermal, Circuit or Johnson noise.
- (ii) Valve noise:-
 - (a) Shot or Shrot noise.
 - (b) Partition or Division noise.
 - (c) Induced Grid noise.

There is good reason in radar for confining attention to this internal noise since at frequencies above about 70 Mc/s man-made static and atmospheric disturbances are both negligible in comparison with unavoidable internal receiver noise. From this point the term noise implies this class of receiver noise, bearing in mind that the use of the word unavoidable does not mean that there can be no improvement, but that with a given design using first class components, ideal contacts, etc., there is a theoretical minimum noise which can be computed for the circuit

4. THERMAL NOISE

In an unenergised conductor the "free" or conduction electrons move about in a random manner, although over any reasonably long time-interval their mass centre remains fixed. At any instant, however, a preponderance of electrons is moving in one direction along the conductor, constituting a minute current whose mean value is zero. This current fluctuation must be associated with a corresponding minute voltage fluctuation across the ends of the conductor,

whose mean value is also zero. To provide a measure of such fluctuations in all cases it is usual to compute the mean square value of the fluctuations about the average value (if any).

It can be shown that the mean square noise current in a circuit consisting of a short-circuited resistor of R ohms is given by:-

$$\overline{i_n^2} = \frac{4}{R} k.T. \Delta f \dots\dots\dots (1)$$

where k is Boltzmann's constant = $1.37 \cdot 10^{-23}$ joules/ $^{\circ}\text{K}$ *

T is the absolute temperature of the resistor in $^{\circ}\text{K}$,
 Δf is the frequency range considered, in cycles/sec, and
 $\overline{i_n^2}$ is in (amps)².

From the above result it follows that the mean square open-circuit noise voltage generated by a resistor of R ohms and measured by an instrument of bandwidth Δf is given by:-

$$\overline{v_n^2} = 4R.k.T \Delta f \dots\dots\dots (2)$$

and if the same units apply, $\overline{v_n^2}$ is in (volts)².

As a general result it can be shown that, for any circuit consisting of ohmic resistors, coils and condensers, if the resistance (dependent on frequency) measured across any two terminals is $R(f)$, then the total noise voltage generated across these terminals is given by:-

$$\overline{v_n^2} = \int_0^{\infty} 4 \cdot R(f) \cdot k.T \cdot df \dots\dots\dots (3)$$

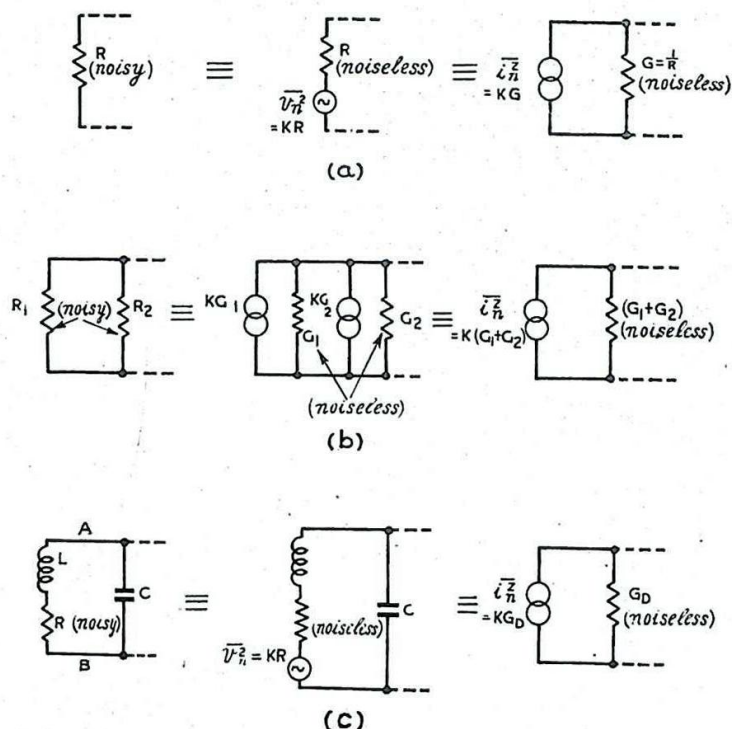


Fig. 643 - Equivalent circuit representations for thermal noise sources.

* $^{\circ}\text{K}$ indicates $^{\circ}\text{Kelvin}$, signifying the use of the Kelvin or Absolute temperature scale, with zero at -273°C .

If the bandwidth accepted by the measuring instrument is limited to Δf , and the resistance can be considered as constant at a value R over this range of frequencies, then:-

$$\overline{v_n^2} = 4RkT\Delta f$$

as obtained above. The circuit representation, in this case, is shown in Fig. 643(a). For convenience we denote $4kT\Delta f$ by K .

When the current-generator equivalent circuit is used, it is sometimes more convenient to work with the conductance G rather than the resistance R of the resistor.

If the circuit consists of two resistors of resistances R_1 and R_2 connected in parallel, the effective short-circuit noise current as measured by an instrument of bandwidth Δf , is given by:-

$$\overline{i_n^2} = K (G_1 + G_2) \dots\dots\dots (4)$$

The equivalent circuit representation is shown in Fig. 643(b).

Consider the case of a high- Q tuned circuit (Fig. 643(c)), for a small bandwidth Δf near resonance. The effective open-circuit noise voltage appearing across AB , due to the resistance R in series with L is given by:-

$$\overline{v_{AB}^2} = Q^2 \cdot \overline{v_n^2} \text{ (where } \overline{v_n^2} = KR)$$

$$\therefore \overline{v_{AB}^2} = Q^2 \cdot K \cdot R.$$

$$\text{But } R_d = \text{Dynamic resistance of tuned circuit} = Q^2 \cdot R = \frac{1}{G_d}$$

$$\therefore \overline{v_{AB}^2} = K/G_d \dots\dots\dots (5)$$

Thus, near resonance, the dynamic resistance of the tuned circuit can be regarded as generating thermal noise. (This is a particular case of the general result quoted above (equation 3)).

It is described in Chap. 17 Sec. 11 how an effective radiation resistance can be attributed to an aerial. This radiation resistance is quite apart from the small, but inevitable, resistance due to the imperfect conductivity of the conductors which constitute the aerial. In calculating the performance of a receiver for amplifying signals from an aerial, the question arises as to what value of noise should be attributed to the radiation resistance of the aerial. It may be shown that if an aerial of radiation resistance R_r is confined in a uniform temperature enclosure, the aerial behaves, for noise considerations, as a resistor of value R_r producing thermal noise fluctuations. That an actual aerial is not in this state is self-evident. However, for lack of a convenient alternative in evaluating the noise in a circuit, it is usual to assign to the aerial a mean square noise voltage of KR_r (where K is normally evaluated at room temperature), in series with the radiation resistance of the aerial. This formula yields results which are sufficiently accurate to be used in practical applications.

SHOT NOISE

5. General

In any valve the emission from the cathode, because it results from the irregular departure of a large number of discrete

units (electrons), has a random nature. Consequently, there are, superimposed on the mean current through the valve, fluctuations which constitute noise. This type of noise is called Shot Noise.

6. Diode Operating Under Temperature-Limited Conditions

The simplest case for consideration, in connection with shot noise, is the temperature-limited diode. The general concept of a valve working under temperature-limited conditions implies that all the electrons after emission proceed immediately to the anode independently of one another. Analysis shows that, with a diode operating under such conditions, the mean square noise current due to the shot effect when anode and cathode are effectively short-circuited is given by:-

$$\overline{i_n^2} = 2 e \cdot I_a \Delta f \dots\dots\dots (6)$$

where e is the magnitude of the electronic charge = $1.59 \cdot 10^{-19}$ coulombs;

I_a is the mean current through the valve in amperes;

Δf is the band-width under consideration,

and $\overline{i_n^2}$ is in (amps)².

7. Diode Operating Under Space-Charge Limitation Conditions

In practice valves are seldom used in their temperature-limited condition. The operating voltages are such that the electrons are not drawn off to the anode immediately after emission from the cathode, but a space charge is built up in the inter-electrode region in the form of an electron cloud.

Under temperature-limitation conditions, and assuming that

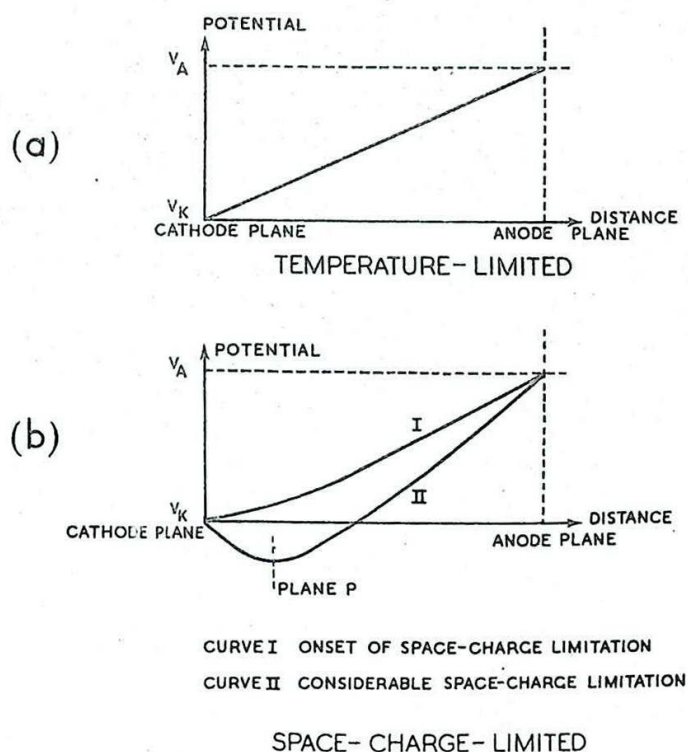


Fig. 644 - Potential distribution in anode-cathode space for a diode (a) temperature limited (b) space-charge limited.

the electrodes are plane and parallel, the potential distribution between the anode and cathode of a diode has the form shown in Fig. 644(a). Under a state of space-charge limitation there is electron interaction in the valve and it may be shown that the potential distribution has a minimum value in the inter-electrode space (Fig. 644(b)). Referring to curve II, near the cathode there is a region in which the field is directed towards the cathode, and between the cathode and anode there is a plane at which the potential is a minimum. Between the potential minimum and the anode, the electric field is directed towards the anode. It follows that, of the electrons emitted from the cathode, only those which have a component of velocity normal to the cathode sufficiently great to carry them past the plane P can ever reach the anode. The remaining electrons do not reach the anode, and thus the total anode current may be considerably less than the emission current.

Consider briefly how space-charge operation affects the shot noise in the current flowing in the anode circuit of the valve. The emission of surplus charge will momentarily increase the space charge and therefore depress the potential minimum. Consequently, certain electrons which would otherwise have had sufficient energy to overcome the potential barrier are now unable to do so. Hence the emission of electrons from the cathode, in any small time-interval δt , at a rate greater than the mean rate of emission, decreases the probability that electrons in immediately succeeding intervals will reach the anode. Thus the electron stream is to some extent ordered by the influence of the space charge, and we should therefore expect the mean square value of the fluctuating component in the anode current to be lower in the case of a valve operating under conditions of space-charge limitation than in one operating under conditions of temperature limitation. The reduction is expressed by a dimensionless factor Γ^2 (or less commonly by F^2 or A) where Γ^2 is known as the Space-Charge Reduction Factor. Under conditions of space-charge limitation, the mean square short-circuit noise current of a diode is given by:-

$$\overline{i_n^2} = 2.e.I_a \Gamma^2 . \Delta f \dots\dots\dots (7)$$

A typical value of Γ^2 , for modern valves working under normal conditions, is of the order of $\frac{1}{20}$.

The value of Γ^2 for a diode depends on the values of the mean valve current I_a , the cathode temperature T_k , which controls the emission current, and the anode slope conductance of the valve ($G_a = 1/R_a$) which gives the ratio of a change of anode current with respect to a change of anode-cathode voltage. A comprehensive analysis gives a relation between Γ^2 and the factors I_a , T_k and G_a which is illustrated by the graph of Fig. 645. In this graph Γ^2 is plotted against:-

$$\frac{I_a \cdot e}{G_a \cdot k \cdot T_k}$$

where e is the magnitude of the electronic charge (1.59×10^{-19} coulombs),

k is Boltzmann's constant (1.37×10^{-23} Joules/°K),

and T_k , G_a and I_a are expressed respectively in °K, mhos and amps. This parameter reduces to:-

$$\frac{11.6 \cdot I_a \cdot 1000}{G_a \cdot T_k} \dots\dots\dots (8)$$

The analysis and the relation illustrated in the figure are valid

only for a relatively high degree of space-charge limitation. However, the condition of a high degree of space-charge is generally realised with modern valves, in particular those with oxide-coated cathodes under normal operating conditions.

An examination of the figure shows that the curve for Γ^2 bears a considerable resemblance to a rectangular hyperbola over the greater part of the range. Assuming that the curve is of this character, it follows that:-

$$\Gamma^2 \cdot \frac{I_a \cdot e}{G_a \cdot k \cdot T_k} = \text{constant.}$$

$$\text{If we write } \theta = \frac{2 \Gamma^2 \cdot I_a \cdot e}{4 G_a \cdot k \cdot T_k}$$

$$\text{we obtain } \Gamma^2 = \frac{4 G_a \cdot k \cdot T_k \cdot \theta}{2 I_a \cdot e} \dots \dots \dots (9)$$

θ has been plotted for various values of $I_a \cdot e$ and Γ^2 in Fig. 645, $G_a \cdot k \cdot T_k$

and it can be seen that over the useful range (i.e. except for very small currents) it is almost ideally constant, with an average value of about 0.65.

The analysis gives Γ^2 and θ for the region between the dotted boundaries of Fig. 645. For the region on the extreme left analysis is complex whilst for the region on the extreme right the condition on which the analysis is based, i.e. a high degree of space-charge limitation, does not hold.

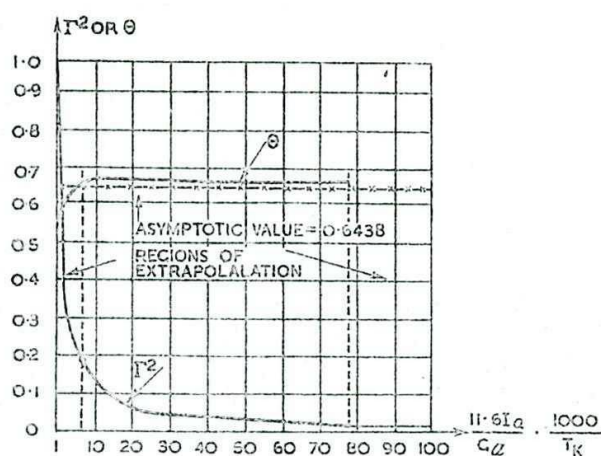


Fig. 645 - Graph of Γ^2 and θ plotted against $\frac{11.6 I_a}{G_a T_k} \cdot 1000$ for a diode operating under conditions of space-charge limitation.

It has previously been stated that the mean-square noise current in a diode working under space-charge limitation conditions is given by:-

$$\overline{i_n^2} = 2e \cdot I_a \cdot \Gamma^2 \cdot \Delta f$$

Replacing Γ^2 by its value obtained from equation (9) we have:-

$$\overline{i_n^2} = 4k (\theta \cdot T_k) G_a \cdot \Delta f \dots \dots \dots (10)$$

This last equation is identical in form with a thermal noise formula, and therefore the shot noise in a diode working under space-charge limitation conditions can be considered formally as thermal noise generated by the diode slope conductance G_a at a modified temperature θT_k .

8. Negative-Grid Triode Under Space-Charge Limitation Conditions

The transition from a diode to a triode may be considered as a process of inserting grid wires at the desired distance from

the cathode. A new space-charge distribution will result which, for sufficiently close grid spacing, will be sensibly uniform over the grid-plane. If, therefore, relatively high- μ valves (close grid spacing) are considered, the analysis for the diode may be extended to triodes. If the grid is sufficiently negative, all electrons which surmount the potential barrier in the cathode-grid space pass through the grid structure to the anode. Thus the current through the grid structure is identical with that reaching the anode. A fictitious collector plate, at some potential, could be placed in the plane of the grid such that it would collect the same space-current as previously passed through the grid structure. If it is feasible to neglect the effect on the space-charge barrier of electrons in the grid-anode space of the triode, then the diode suggested above would, for noise considerations, be a truly equivalent diode. For a finely-wound grid structure, this hypothesis is admissible, since not only are the electrons in the grid-anode space further from the space-charge barrier than those in the cathode-grid space, but also the grid acts as an electrostatic screen between the electrons in the grid-anode space and the space-charge barrier.

The problem resolves itself into a determination of the relationship between the conductance of this equivalent diode, and some measurable quantity relating to the actual triode. It is to be expected that the slope conductance of the equivalent diode G_a is related to the mutual conductance of the triode G_m and simple analysis shows in fact that the following relationship holds.

$$\frac{G_m}{\sigma} = G_a \dots\dots\dots (11)$$

where σ is a function of amplification factor, transit-time factors and electrode spacings of the triode. If the amplification factor is very large σ is almost unity. For all conventional valves σ lies between 0.5 and unity, and is nearly unity for a pentode.

Using the results expressed in equations (10) and (11), we see that the mean square noise current in a negative-grid triode, operating under space-charge limitation conditions, is given by:-

$$\overline{i_n^2} = \frac{\theta}{\sigma} \cdot 4 \cdot k \cdot T_k \cdot G_m \cdot \Delta f \dots\dots\dots (12)$$

From equations (9) and (11)

$$\tau^2 = \frac{2 \cdot k \cdot T_k \cdot \theta \cdot G_m}{I_a \cdot e \cdot \sigma}$$

or including the numerical values of k , e and θ

$$\tau^2 = \frac{0.11}{\sigma} \cdot \frac{G_m}{I_a} \cdot \frac{T_k}{1000} \dots\dots (13)$$

Fig. 646 shows the variation of I_a , G_m and τ^2 for an Acorn Triode (Mullard AT4). The anode voltage is 200 volts and σ has been taken as unity.

9. PARTITION NOISE

Since the current in a valve is composed of discrete entities emitted from the cathode

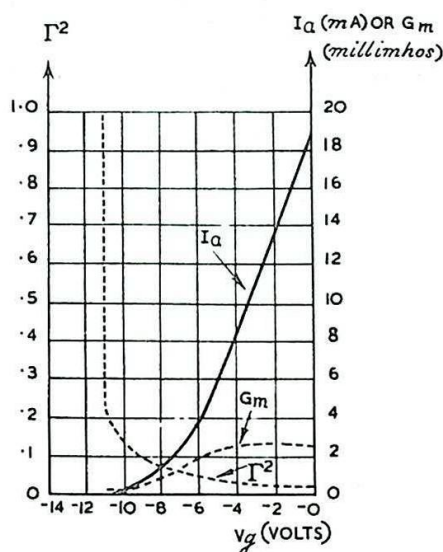


Fig. 646 - Graph showing values of I_a , G_m and τ^2 for a Mullard AT4 for different values of grid-cathode voltage.

with random direction and velocity the precise distribution of the electron stream at any instant is unpredictable. In the case of a diode or negative-grid triode, it can be said that all electrons which surmount the space-charge barrier will arrive at the anode. In the case, however, of a valve with more than one collecting electrode no more can be said than that, as the result of the transit of a large number of electrons, the mean currents to the anode and the other collecting electrode are both known. The final destination of each individual electron will be a matter of pure chance. This form of randomness relating to the valve current will constitute a fresh source of noise, which is known as Partition Noise.

It is to be expected, on the basis of random division of the electron stream of a valve between two collectors, the anode and screen-grid, that the partition noise in the anode circuit is equal to that in the screen circuit. Further, the partition noise should depend only on the total valve current and the relative division of anode and screen currents. At first sight it might seem a formidable problem to confirm these suppositions experimentally or even to show the existence of partition noise independently of shot noise. However, a relatively simple experiment achieves the object. In the circuit shown in Fig. 647, if the cathode resistor R is made sufficiently large the equivalent mutual conductance G_m can be reduced to a negligibly small value while the mean conduction current I_a is maintained by suitable electrode potentials. It has been shown (equation (13)) that the shot noise reduction factor Γ^2 is proportional to

$$\frac{G_m}{I_a} ;$$

hence a sufficient reduction of G_m reduces the anode shot noise

($i_n^2 = 2eI_a\Gamma^2 \Delta f$) to negligible

proportions. The partition noise, on the other hand, providing it

depends only on I_a and the relative division of anode and screen current, should remain unchanged, and in fact does so.

In the entire absence of shot noise (a purely theoretical condition) the mean-square short-circuit partition noise current in the anode (or screen) circuit is given by:-

$$i_n^2 = 2e \cdot \frac{I_a \cdot I_s}{I_a + I_s} \cdot \Delta f \dots\dots\dots (14)$$

where i_n^2 is measured in (amps)²,

and I_a and I_s are respectively the mean anode and screen currents in amps.

It is evident that the partition noise in the anode circuit of a pentode decreases to zero as the mean current flowing to the screen grid is reduced to zero. Hence design which results in the reduction of the mean screen current leads to less noisy pentodes. The screen current can be reduced in one of two ways:-

- (i) The pitch of the screen grid can be increased and the

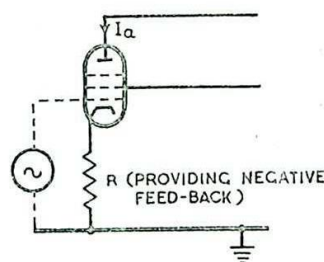


Fig. 647 - Circuit for reducing shot noise without affecting partition noise.

diameter of its wire reduced. This, however, reduces the screening between anode and grid, and decreases the control of the screen grid on the electron stream.

- (ii) The method of beaming can be used. This can be achieved by arranging the meshes of the various grids in such a way that the electron stream passes between the wires of the screen grid, i.e. the electron stream is focused.

In general shot and partition noise are present together. An analysis which considers the contributions to noise in the anode lead by electrons which:-

- (i) Cross the potential barrier, which is due to space charge, and finally reach the anode circuit,
- (ii) Cross the potential barrier and finally reach the screen grid circuit,
- (iii) Do not have sufficient emission energy to cross the potential barrier, but which contribute to anode and screen circuit noise through resultant fluctuations of the potential barrier,

leads to the general formula:-

$$\overline{i_n^2} = 2eI_a \cdot \left(\frac{I_s + \Gamma^2 I_a}{I_a + I_s} \right) \cdot \Delta f \dots\dots\dots (15)$$

In the formula (15) $\overline{i_n^2}$ represents the total mean square noise current in the anode circuit, and Γ^2 is computed on the basis of the total current I in the valve. It is clear that, if $\Gamma^2 = 0$ (i.e. theoretical absence of shot noise), expression (15) reduces to expression (16).

Examples of the value of the mean square noise current in the anode circuit of a valve

We shall evaluate the mean square noise currents in the anode circuit of a valve, which is operated firstly as a pentode and secondly as a triode.

The following data applies to an Acorn Pentode (Mullard AP4) operating as a pentode:-

$$G_m = 2.25 \text{ mA/volt} : I_a = 6 \text{ mA} : I_s = 1.8 \text{ mA} : I = I_a + I_s = 7.8 \text{ mA}.$$

The temperature T_k of the cathode of the valve can be taken as 1,100°K.

The mean square noise current in the anode circuit of the pentode is given by:-

$$\overline{i_n^2} = 2e \cdot I_a \left(\frac{I_s + \Gamma^2 \cdot I_a}{I_a + I_s} \right) \Delta f \text{ (from expression (15)).}$$

$$\text{where } \Gamma^2 = \frac{0.11}{\sigma} \cdot \frac{G'_m}{I} \cdot \frac{T_k}{1000}.$$

As an approximate relation

$$\frac{G_m}{G'_m} = \frac{I_a}{I} \text{ or } \frac{G'_m}{I} = \frac{G_m}{I_a}$$

where G'_m is the total conductance of the pentode, $\frac{\partial I}{\partial V_g}$.

Assuming $\sigma = 1$

$$\begin{aligned}\tau^2 &= 0.11 \cdot \frac{2.25}{6} \cdot \frac{1100}{1000} \\ &= 0.045.\end{aligned}$$

Consequently

$$\begin{aligned}\overline{i_{pn}^2} &= 2e \cdot \Delta f \cdot 6 \cdot 10^{-3} \cdot \left(\frac{1.8 + 0.045 \cdot 6}{7.8} \right) \\ &= 2e \cdot \Delta f \cdot 1.60 \cdot 10^{-3}.\end{aligned}$$

The mean square noise current in the anode circuit of the valve used as a triode is given by:-

$$\overline{i_{tn}^2} = 2e I \tau^2 \Delta f.$$

Assuming that the total space current is the same as that when the valve is used as a pentode

$$\begin{aligned}\overline{i_{tn}^2} &= 2e \cdot \Delta f (7.8 \cdot 0.045) \cdot 10^{-3} \\ &= 2e \cdot \Delta f \cdot 0.35 \cdot 10^{-3}.\end{aligned}$$

It follows that the valve used as a triode is less than one quarter as noisy as it is when used as a pentode.

10. INDUCED GRID NOISE

The primary noise sources (thermal, shot and partition) discussed in the previous sections, have a common characteristic; the mean square noise current is proportional to the total bandwidth involved, but is otherwise independent (at least over a wide range of frequencies) of frequency.

Another type of noise known as Induced Grid Noise (sometimes as High Frequency Noise) results from the movement of charges in the grid circuit which are induced electrically as a result of variations in the space current. If an external impedance is present between grid and cathode a noise voltage is generated which will, in turn, react upon the electron stream in the valve to produce additional valve noise.

The mean square noise current due to induced grid noise

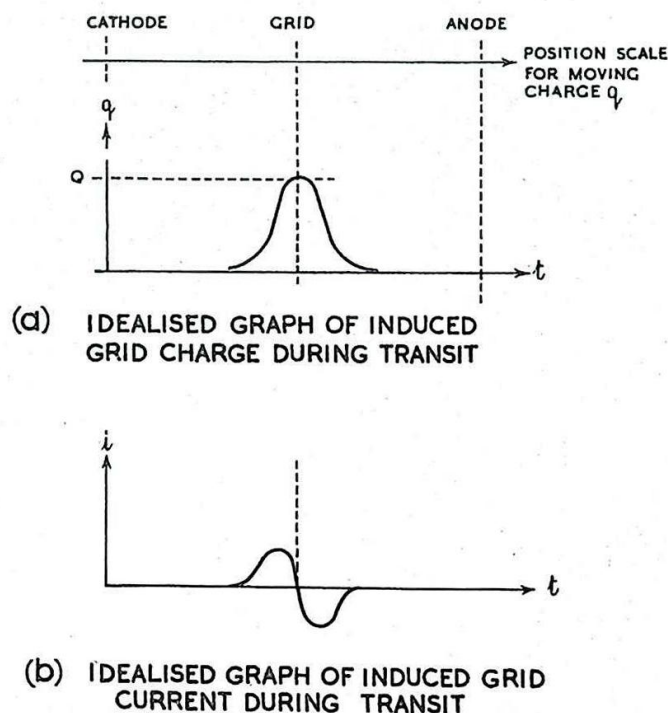


Fig. 648 - Induced charge and current at grid.

is dependent on frequency and is in fact of importance only above the frequency limit where transit-time effects become significant.

The process leading to the production of induced grid noise can be explained as follows:-

If there is a uniform flow of electrons through a valve any induced current in an earthed grid, brought about by the approach of electrons to the grid, is counteracted by an opposite induced current produced by electrons receding from the grid. However, if a concentration of charge $-Q$, surplus to the average flow of charge in the anode-cathode space, passes across the valve and through a closely-wound grid mesh, then the variation with time of the charge q induced on the grid is of the form shown in Fig. 648(a). Consequently, since the induced current is the rate of change of the induced charge with time, the current pulse in the external grid circuit takes the form shown at (b).

If a Fourier analysis is performed with respect to a pulse of this form, it may be shown that no appreciable contribution of energy is made by frequency components small compared with $\frac{1}{\tau}$ where τ is the time of transit of electrons through the valve. Thus high frequency noise is evident only at frequencies for which there is a significant component of input conductance due to transit time (Chap. 7 Sec. 25).

Analysis gives the following expression for the mean square short-circuit noise current in the grid circuit:-

$$\overline{i_n^2} = 4k.T. 4.8 G_t . \Delta f \dots\dots\dots (16)$$

where T is the room temperature ($^{\circ}K$).

k is Boltzmann's constant

G_t is the input conductance (mhos) due to transit-time loading only

and $\overline{i_n^2}$ is in (amps)².

These formulae are valid for only the initial part of the frequency range in which transit time is important.

The equivalent circuit for computation of the mean square grid fluctuation voltage $\overline{v_n^2}$ is shown in Fig. 649. The component of mean square short-circuit noise current in the anode circuit is then given by:-

$$\overline{i_n^2} = G_m^2 . \overline{v_n^2} \dots\dots\dots (17)$$

Induced grid noise, which is noise produced in the circuit of an electrode not a collector of electrons, is of considerable importance in UHF in such valves as the klystron (see Chap. 16 Sec.1).

Example

A valve operating at 200 Mc/s has an input conductance G_t due to transit time of 0.0005 mhos. The external circuit conductance G (including leakance) is 0.0005 mhos. The mutual conductance G_m of the valve is 0.003 mhos. The bandwidth Δf of the receiver is 2 Mc/s and room temperature T is 300 $^{\circ}K$. Find the contribution of induced grid noise to the short-circuit noise current in the anode circuit.

$$\overline{g_{in}^2} = 4 \cdot k \cdot T \cdot 4 \cdot 8 \cdot G_t \cdot \Delta f$$

from expression (16).

Therefore

$$\overline{g_{in}^2} = 8 \cdot 10^{-17} \text{ (amps)}^2$$

$$\overline{g_{vn}^2} = \frac{\overline{g_{in}^2}}{(G + G_t)^2} \text{ (See Fig. 8)}$$

$$= \frac{8 \cdot 10^{-17}}{(2 \cdot 0.0005)^2} = 8 \cdot 10^{-11} \text{ (volts)}^2.$$

$$\overline{a_{in}^2} = G_m^2 \cdot \overline{g_{vn}^2} \text{ from (17)}$$

$$= \frac{9}{10^6} \cdot 8 \cdot 10^{-11}$$

$$= 72 \cdot 10^{-17} \text{ (amps)}^2$$

$$\text{Thus } \sqrt{\overline{a_{in}^2}} = 2.68 \cdot 10^{-8} \text{ RMS amps}$$

$$= 0.0268 \text{ RMS microamps.}$$

If the total effective output resistance is 10,000 ohms, the noise voltage developed in the output is given by:-

$$\overline{a_{vn}^2} = 0.0268 \cdot 10,000 = 268 \text{ RMS microvolts.}$$

11. FURTHER SOURCES OF NOISE

Apart from the shot, partition and induced grid noise in valves, thermal noise in the resistors and noises generated in the aerial, which have already been described, there are various other sources of noise which may arise in circuits. These are briefly mentioned here, for completeness, although they have very little practical importance in radar circuits.

(i) Noise arising from Flicker Effect

This is a type of noise which sometimes occurs in audio-frequency amplifiers, particularly if valves with oxide-coated cathodes are used. It is probably due to random changes of the emitting properties of the various parts of the cathode. The duration of these changes is of the order of a millisecond. Owing to this relatively long duration of the fluctuations, the noise voltage produced across an anode load is large only if the load is of high impedance for frequencies below about 1 Kc/s. Consequently Flicker effect is of small importance at radio-frequencies.

(ii) Noise due to Ionisation Effects

If a valve is imperfectly evacuated, and ionisation takes place, there is a further source of noise. This is of little practical significance since a high vacuum normally obtains in modern valves.

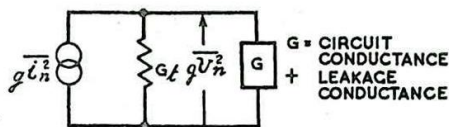


Fig. 649 - Equivalent circuit for computation of mean square voltage in grid circuit due to induced grid noise.

(iii) Secondary Emission Effects

For some operating conditions of a valve, appreciable secondary emission of electrons may take place from an electrode bombarded by electrons from the cathode. This secondary emission results in an increase in the noise of the valve.

(iv) Noise arising from certain types of resistors

Certain non-metallic resistors (carbon resistors) may give rise to fluctuation voltages of a random nature when a direct current passes through them. This noise, which is in excess of thermal agitation, is proportional to current and also increases with the resistance.

THE EQUIVALENT NOISE RESISTANCE OF A VALVE

12. General

In the computation of total noise in a receiver and in particular of the noise resulting from a single amplifier stage it is necessary to combine the effect of thermal noise in the grid circuit with that due to shot and partition noise arising in the valve. It is desirable to replace noise sources in various parts of the circuit by a single equivalent noise generator at some suitable point; this is called "referring" the noise to this point. To simplify calculation and provide rapid comparison of the merits of different valves from the aspect of signal-to-noise performance it is most convenient to refer valve noise to the grid circuit.

In the grid circuit is placed a fictitious noise generator which would produce in an ideal noiseless valve a quantity of noise equal to that actually produced in the valve under examination. To facilitate immediate combination with true thermal noise generated in the grid circuit by ohmic resistance present, it is standard practice to represent the valve noise generator as a fictitious noisy resistance in series with the external grid circuit. This resistance "generates" the requisite amount of thermal noise to account for the valve noise present. Thereafter the valve itself may be regarded, for the purpose of calculation, as an ideal noiseless amplifier.

13. Triode Valves

In Fig. 650, R_e is the fictitious noise resistance to be computed. The mean square short-circuit shot noise current in the anode circuit of a triode is given by:-

$$\overline{i_n^2} = \frac{e}{\sigma} \cdot 4 \cdot k \cdot T_k \cdot G_m \cdot \Delta f.$$

(see expression (12)).

The mean square noise voltage generated by R_e is given by:-

$$\overline{v_n^2} = 4 \cdot k \cdot T \cdot R_e \cdot \Delta f = K \cdot R_e$$

(see expression (2))

and the resulting mean square

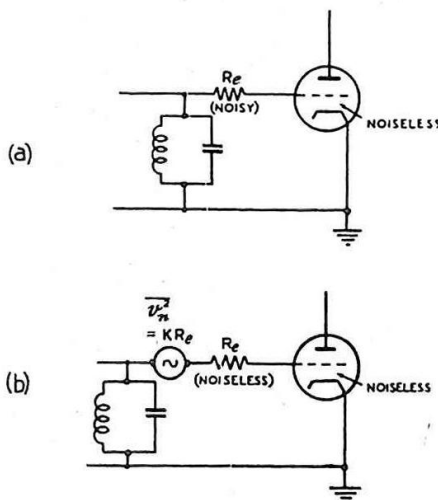


Fig. 650 - Circuit illustrating equivalent shot noise resistance of a valve.

short-circuit noise current in the anode circuit is:-

$$\begin{aligned} \overline{i_n^2} &= G_m^2 \cdot \overline{v_n^2} \\ &= G_m^2 \cdot 4 \cdot k \cdot T \cdot R_e \cdot \Delta f. \end{aligned}$$

Hence it follows that:-

$$G_m^2 \cdot 4 \cdot k \cdot T \cdot R_e \cdot \Delta f = \frac{\theta}{\sigma} \cdot 4 \cdot k \cdot T_k \cdot G_m \cdot \Delta f$$

$$\text{or} \quad R_e = \frac{T_k}{T} \cdot \frac{\theta}{\sigma} \cdot \frac{1}{G_m} \dots\dots\dots (18)$$

It will be noted that R_e is independent of the bandwidth considered and this is an important advantage of this method of estimating valve noise. In the case of valves with oxide-coated cathodes, assuming

$$T_k = 1,100^\circ\text{K}, T = 300^\circ\text{K} \text{ and } \theta = 0.65:-$$

$$R_e = \frac{2.4}{\sigma} \cdot \frac{1}{G_m} \dots\dots\dots (19)$$

A commonly quoted figure for R_e is $\frac{2.5}{G_m}$ and this formula always gives a reasonable estimate of the equivalent noise resistance referred to the grid circuit.

14. Pentode Valves

For a pentode valve it is usual to combine shot and partition noise in the formula for the equivalent noise resistance. This resistance is then given by:-

$$R_e = (1 + 8.7 \sigma \frac{I_s}{G_m} \frac{1000}{T_k}) (\frac{2.5}{\sigma} \cdot \frac{I_a}{I_a + I_s} \cdot \frac{1}{G_m})$$

G_m being the anode-current, grid-voltage mutual conductance, $\frac{\partial I_a}{\partial v_g}$.

The following table gives the noise resistances of some common valves operating under specified conditions.

Valve Type	Valve Class	Operating Conditions			
		$I_a(\text{mA})$	$I_s(\text{mA})$	$G_m(\text{mA/volt})$	R_e
AP4	Pentode (Acorn)	6	1.8	2.25	6,120
6SK7 (CV1981)	Pentode	9.2	2.4	2	10,500
6SJ7 (CV591)	Pentode	3	0.8	1.6	5,800
6AC7 (CV660)	Pentode	10	2.5	9	720
RCA956 (CV649)	Pentode	5.5	1.8	1.8	9,400
RCA955 (CV1059)	Triode	4.5	-	2.0	1,250

It should be noted that the expressions quoted above for the pentode valve are based on two plausible assumptions concerning the behaviour of electrons emitted from the cathode:-

(i) that the "fate" of every electron emitted (i.e. whether it terminates on screen or anode) is a matter of pure chance independent of its point of emission from the cathode.

(ii) that the influence of an individual electron may be regarded as "spread out" over the whole space-charge. In other words, the depression of the space-charge minimum as a result of "surplus" electrons (see Sec.7) is assumed to act in every case on the anode and screen currents in the same manner, in proportion to their mean values I_a , I_{sg} .

These assumptions may not be fully justified in modern valves and therefore in some cases a small correction factor may be found desirable in practice to allow, for example, for the effect of "grid alignment", which would tend to alter assumption (i) above. Again, very close electrode spacing without reduction of the cross-sectional area of the cathode might result in violation of assumption (ii) calling for suitable modification.

15. VALVE NOISE UNDER FEEDBACK CONDITIONS

If negative feedback is present in a valve amplifier the noise current generated in the output circuit is modified since the noise currents flowing in the feedback circuit will introduce compensating voltages between grid and cathode which will reduce the noise. This is illustrated in Fig.651 for the important case of current feedback (see Chap.16 Sec.12). Fundamentally, the effect of the feedback on the noise current may be considered as the result of a limiting process: each component of noise current produces a corresponding feedback voltage at the input which in turn causes a further (diminished) noise current to flow in the valve and so on. The result of adding the various terms of the series thus formed, for the circuit of Fig.651, in which the anode of the valve is effectively short-circuited to earth, is to reduce the noise current which flows through the feedback impedance according to the formula

$$(\overline{i_n^2})_{\text{with feedback}} = \frac{(\overline{i_n^2})_{\text{without feedback}}}{(1 + R_f G_m)^2} \quad (\text{compare Chap. 7 Sec.16}).$$

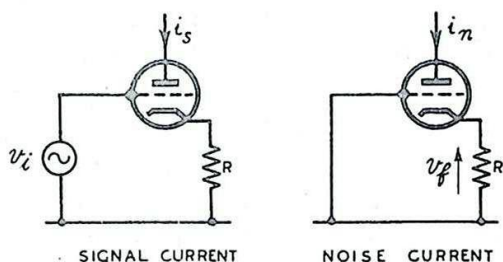
If now we consider the formula (15) for noise in a pentode viz,

$$\overline{i_n^2} = 2e I_a \frac{(I_s + \Gamma^2 I_a) \Delta f}{I_s + I_a}$$

then it is the first term

$$\frac{2e I_a I_s \Delta f}{I_a + I_s}, \text{ depending only}$$

on the ratio of anode and screen currents, which is unaffected by feedback, whereas the second term, corresponding to the proportion of shot noise developed in the output circuit, is affected by the feedback as indicated above. The presence of the Γ^2 factor in this term,



(The steady component of current, I_a , is omitted from the diagrams)

Fig. 651 - Noise under feedback conditions.

$$\frac{2e I_a^2 r^2 \Delta f}{I_a + I_s}$$

indicates the susceptibility to feedback, since it expresses the dependence of the noise current on the grid-cathode voltage and space-charge. On the other hand, the product $I_a I_s$ is, to a first approximation, independent of variations in the cathode field and is therefore independent of the feedback.

16. SUMMARY

It has now been shown in the preceding sections that the physical explanation for the requirement of a comparatively high signal-to-noise ratio at the input stage of a receiver lies in the fact that while there are no new primary sources of signal power in the receiver between input and display there are many new sources of noise of different kinds, and the signal-to-noise ratio must inevitably deteriorate (if the same bandwidth is maintained) from aerial to output.

In the PPI type of display the signal and noise appear as variations in the brightness of the trace. Under such circumstances it is correspondingly more difficult to distinguish a weak signal from the noise background than with an A-Type display, and therefore the signal-to-noise ratio just prior to the detector of the receiver must be greater than unity (perhaps greater than 4) if the signal indication is to be distinguished from the noise.

On the other hand when the signal is used to measure angles of elevation or azimuth it is possible to use the integrated effect of the gated signals from successive recurrence periods so that the bandwidth of the signal fed to the output indicator or automatic following unit is very much less than the bandwidth at the detector stage immediately following the video frequency amplifier. This reduction in bandwidth may result in a considerable improvement in the resultant signal-to-noise ratio at the output, since the required signal power is concentrated in a relatively narrow band and is not reduced in the same ratio as the noise power, which is distributed uniformly throughout the band. For this reason it may be possible for a radar set to track successfully by auto-following a target which does not give a signal-to-noise ratio of even 1:1 at the detector stage.

CHAPTER 16

RECEPTION OF RADIO FREQUENCY PULSES

INTRODUCTION

1. General

In radar equipments the transmitted RF pulse possesses a more or less rectangular waveform; (Fig. 652). After transmission and reflection from the target the pulse returned to the equipment has approximately the same shape but is very much weaker. The ideal receiver should be capable of receiving the weakest possible reflected pulse without distorting its shape or introducing any noise.

In Chap. 15 it has been shown that unavoidable noise power is generated in any receiver and it is this noise power which ultimately limits the sensitivity and fidelity of pulse reception. A compromise design must be adopted and it is therefore inevitable that careful attention is given to the problem of noise when radar receiver design is under consideration.

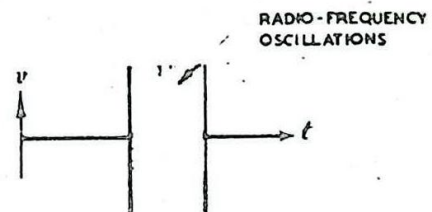


Fig. 652 - Radio-frequency pulse of rectangular waveform.

Clearly the maximum possible sensitivity should be secured in the receiver, anything less than this wasting transmitted power, and it is therefore usual to find that radar receivers have sufficient maximum gain to produce saturation or near-saturation outputs from inherent noise alone.

The distortion of the RF pulse in the receiver is important for two different reasons :-

- (i) If the Range difference to two separate targets corresponds to a time interval greater than half the duration of the transmitted pulse, then the target echoes are received separately ; otherwise the two responses partially coincide. The resolving power in distance is therefore limited by the duration of the emitted pulse, and any increase of pulse length due to receiver distortion will reduce this resolving power, i.e. there will be a loss in Range discrimination.
- (ii) An attempt is made to keep the leading edge of the reproduced pulse as abrupt as possible, particularly when measurement of this leading edge is used for the determination of Range. Any loss of steepness of this edge due to distortion in the receiver results in a corresponding loss of accuracy of Range measurement.

In a radar equipment, the direct signal from the nearby transmitter produces extremely large voltages in the receiver. These voltages, which may persist for a time considerably longer than the transmitted pulse, usually render the receiver insensitive to signals. This effect is known as Paralysis, and unless it is minimised or overcome the receiver will be incapable of responding adequately to signals from targets at close range.

The receivers in normal use in radar employ the Superheterodyne principle. An arrangement, shown in Fig. 653, consists of RF Amplifiers,

a Mixer (Frequency Changer) and Local Oscillator, IF (Intermediate Frequency) Amplifiers, a Detector, and Video Frequency Amplifiers. The radio frequencies in normal use in radar lie in the range 30 Mc/s. to 10,000 Mc/s. It has not yet been found possible to produce useful amplification at frequencies greater than about 700 Mc/s.; hence, in Radar receivers operating at frequencies greater than this, the RF Amplifiers are omitted. Idealised waveforms of the pulse to be expected at different stages of the receiver are shown in Fig. 654.

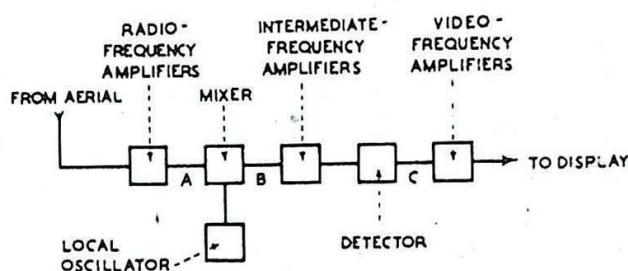


Fig. 653 - Block diagram of superheterodyne receiver.

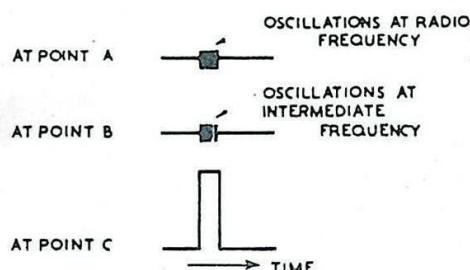


Fig. 654 - Idealised waveforms of pulse at various points of the superheterodyne receiver shown in fig. 653.

2. Fourier Analysis

The signal which is transmitted and received by Radar equipments is ideally a rectangular-shaped RF pulse (Fig. 652) of short duration ($1/10$ to 10 microseconds). The Bandwidth of the receiver which is required to receive such pulses without undue distortion can be calculated.

Consider the succession of pulses shown in Fig. 655. If these are exactly similar in shape and are repeated at regular intervals of time ($\frac{1}{f}$) they can be analysed as a Fourier spectrum consisting of a number of sinusoidal components of different amplitude, frequency and phase. In the case under consideration, assuming that one pulse is located symmetrically about an axis representing zero time, the components have a cosine variation, i.e., the instantaneous value of each is a maximum at zero time. The Fourier series corresponding to the waveform of Fig. 655 is thus :-

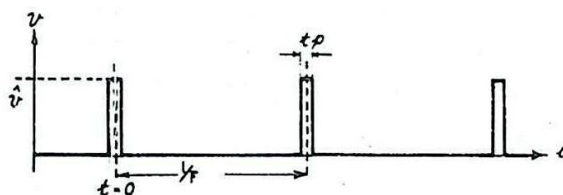


Fig. 655 - Succession of rectangular pulses.

$$v = a_0 + a_1 \cos \omega t + a_2 \cos 2 \omega t + a_3 \cos 3 \omega t + \dots \text{to infinity}$$

where $\omega = 2 \pi F$ (1)

There is thus a steady component of value a_0 , a component of amplitude a_1 and of frequency equal to the repetition frequency of the pulses (fundamental component, or first harmonic), together with an infinite number of components which are harmonics of this frequency.

If each pulse is of amplitude \hat{v} and duration t_p , calculation shows that the value of the steady component is $\hat{v} F t_p$, the amplitude a_1 of the fundamental is $\frac{2\hat{v}}{\pi} \sin \pi F t_p$ and the amplitude a_n of the n th harmonic

$$\frac{2\hat{v}}{n\pi} \sin n \pi F t_p.$$

harmonics of such an order that $n \pi F t_p$ is equal to a multiple of π have zero amplitude. These components have frequencies $\frac{1}{t_p}$, $\frac{2}{t_p}$, $\frac{3}{t_p}$, etc., and their positions on the amplitude spectrum are known

respectively as the first, second, third, etc., zeros. A typical amplitude spectrum, for pulses of one volt amplitude, one microsecond duration and of repetition frequency 400 per second is shown, as far as the third zero, in Fig. 656. In theory the spectrum covers an infinite range of frequencies. Since $t_p = 1$ microsecond the zeros are spaced 1 Mc/s. apart, the individual components being separated by 400 c/s., i.e., by the repetition frequency of the pulses. There are thus 2,500 components between zero frequency and the first zero of the spectrum, so that a large scale is required if the individual components are to be represented, as they should be, by discrete lines.

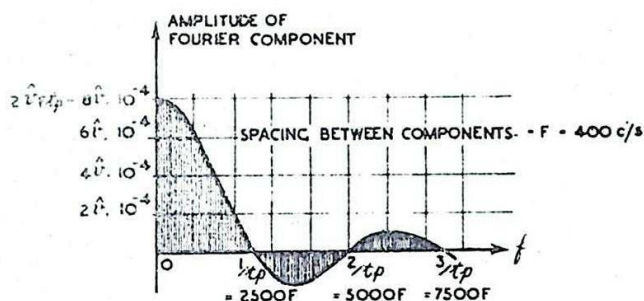


Fig. 656 - Amplitude spectrum of recurrent rectangular pulses with a repetition period = 2500 X pulse duration.

It is shown in the Standard Service Manuals (see BR 230

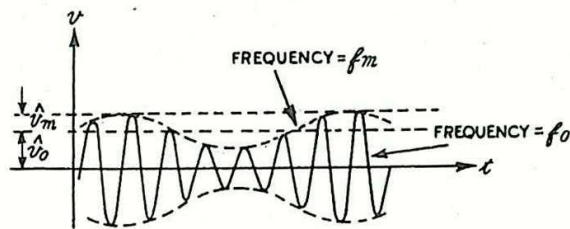
(Admiralty Handbook of Wireless Telegraphy, Vol II) Sec. N para. 15, and AP 1093, Chap. XII para. 14 for a more detailed analysis) that a continuous sinusoidal oscillation of frequency f_0 , amplitude modulated at a frequency f_m , behaves as though it were composed of three sinusoidal components of constant amplitude and of frequencies f_0 , $f_0 + f_m$ and $f_0 - f_m$ respectively. This simple example is illustrated in Fig. 657, and the equation of the modulated wave is :-

$$v = (\hat{v}_0 + \hat{v}_m \cos \omega_m t) \cos \omega_0 t$$

$$= \hat{v}_0 \cos \omega_0 t + \frac{1}{2} \hat{v}_m \cos (\omega_0 + \omega_m) t + \frac{1}{2} \hat{v}_m \cos (\omega_0 - \omega_m) t$$

where $\omega_0 = 2 \pi f_0$ and $\omega_m = 2 \pi f_m$ (2)

The three components are known as the carrier, upper sideband and lower sideband respectively. The amplitude spectrum is shown in Fig. 658. It should be noted that the carrier amplitude \hat{v}_0 is equal to the average value of the modulated carrier, and the amplitudes of the sidebands are each equal to half the amplitude of the modulating oscillation.



A CONTINUOUS OSCILLATION WHICH IS AMPLITUDE - MODULATED.

Fig. 657 - A continuous oscillation which is amplitude-modulated.

When the modulation envelope is not a pure sine wave it may be analysed into a Fourier series and each harmonic in this series will give rise to its own pair of sidebands. The greater the number of harmonics in the modulating waveform the more numerous are the sidebands and the wider is the spread of the frequency spectrum on each side of the carrier.

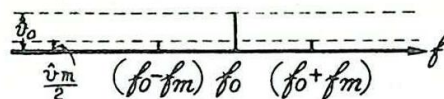


Fig. 658 - Amplitude spectrum of the amplitude-modulated oscillation shown in fig. 657.

It has already been shown that a succession of pulses (Fig. 655) can be considered as consisting of sinusoidal components (Equation (1)). Hence if the successive RF pulses shown in Fig. 659 are

exactly similar in shape, are repeated at regular intervals and are Coherent, i.e., they are the result of amplitude modulation of a continuous sinusoidal oscillation, the equation representing these pulses will be :-

$$v = \cos \omega_0 t (a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + a_3 \cos 3\omega t + \dots)$$

where $\omega = 2\pi F$

$$= a_0 \cos \omega_0 t \dots \dots \dots (\text{carrier})$$

$$+ \frac{1}{2} a_1 \cos (\omega_0 + \omega)t + \frac{1}{2} a_1 \cos (\omega_0 - \omega)t \quad (\text{first sidebands})$$

$$+ \frac{1}{2} a_2 \cos (\omega_0 + 2\omega)t + \frac{1}{2} a_2 \cos (\omega_0 - 2\omega)t \quad (\text{second sidebands})$$

$$+ \frac{1}{2} a_3 \cos (\omega_0 + 3\omega)t + \frac{1}{2} a_3 \cos (\omega_0 - 3\omega)t \quad (\text{third sidebands})$$

$$+ \text{etc.} \quad \dots \dots \dots (3)$$

The carrier amplitude is therefore $\hat{v} F t_p$ and the sideband amplitudes $\frac{\hat{v}}{\pi} \sin \pi F t_p$, $\frac{\hat{v}}{2\pi} \sin 2\pi F t_p$, $\frac{\hat{v}}{3\pi} \sin 3\pi F t_p$, etc. If the pulse

recurrence frequency is 400/sec. and the pulse duration 1 microsecond as before, the amplitude spectrum for $\hat{v} = 1$ volt and $f_0 = 500$ Mc/s. is shown in Fig. 660.

The changes which take place in the amplitude spectrum in the transition from the recurrent pulses of Fig. 655 to the recurrent coherent

RF pulses of Fig. 659 may now be seen. The steady component of magnitude $\sqrt{F}t_p$ becomes the carrier component having the same peak amplitude, whilst each of the harmonic components of Fig. 656 splits into a pair of sidebands of half the amplitude and symmetrically disposed with respect to the carrier. Given Fig. 656 it is thus possible to derive Fig. 660 by this simple process.

In practice the RF pulses received by radar equipments are not repeated at exactly equal intervals of time, and successive pulses are likely to be slightly different in duration and amplitude. Also such pulses are usually not coherent in the sense described above; i.e., the radio-frequency oscillation starts afresh in each pulse and there is no phase linkage from one pulse to the next. The effect of these variations in duration, amplitude and radio-frequency phase is to "blur" the spectrum so that it no longer consists of discrete frequency components, without, however, altering its general nature.

It has been shown that the periodic train of pulses shown in Fig. 655 consists of sinusoidal components of frequencies F , $2F$, $3F$, etc. As the repetition frequency is reduced the components thus become correspondingly closer in frequency; at the same time the amplitude of each is reduced. In the limiting case of a single pulse the components become indefinitely close together, thus forming a continuous spectrum, while the amplitudes are vanishingly small. However it may be shown that the relative amplitudes of the components are still given by the envelope of Fig. 656. This argument applies equally well to the RF pulses of Fig. 659 and is the basis of the Fourier Integral treatment of the single RF pulse.

In radar, single RF pulses are not used, the recurrence frequency seldom being less than 400 c/s. However, the amplitude-frequency spectrum is more accurately portrayed as a continuous or "White" spectrum rather than one composed of discrete harmonic components, for the following reasons :-

- (i) As already remarked, the pulses are not coherent, there being no phase linkage between successive transmitted pulses.
- (ii) The transmitter frequency is liable to vary by an amount sufficient to blur the spectrum completely.

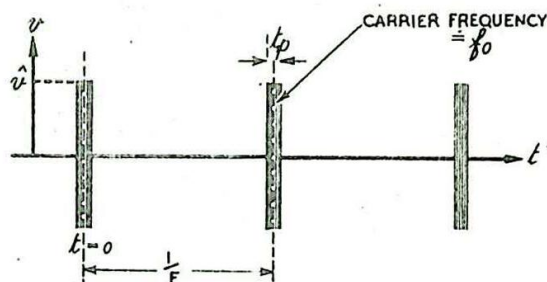


Fig. 659 - Recurrent radio-frequency pulses.

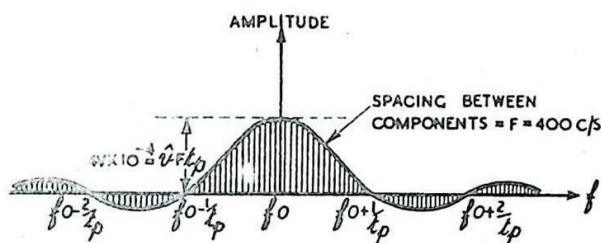


Fig. 660 - Amplitude spectrum of recurrent RF pulses with a repetition period = $2500 \times$ pulse duration.

E.G., suppose the recurrence frequency is 1000 c/s. For a centimetre equipment a random frequency-variation of 1000 c/s between successive pulses is not unlikely. Hence the positions of the harmonic components of the spectrum are completely-random.

- (iii) The pulse width t_p and recurrence frequency F , and therefore the envelope of the amplitude-frequency spectrum, are liable to random variations.

3. Distortion of RF Pulse During Amplification

As described in Chap. 7 Sec. 8, in order to amplify satisfactorily RF pulses of the type shown in Fig. 652 an RF amplifier must exhibit amplitude-frequency and phase-frequency characteristics of the types shown in Fig. 661. These characteristics are drawn in an idealised form in Fig. 661, and a comparison with Fig. 660 shows that with such characteristics the higher order sidebands are inevitably removed in the amplifier. The wider the pass band, the more accurately is the pulse envelope reproduced; but at the same time the noise is increased, and it is therefore necessary to adopt a compromise, the exact nature of which depends on the requirements of the particular radar application.

In the first instance we may calculate the response of an idealised band-pass amplifier, possessing characteristics such as those shown in Fig. 661, to a succession of coherent RF pulses. The response of such an amplifier to one of these pulses is illustrated in Fig. 662.

The input pulse is shown at (a), while (b) and (c) show the shape of the output pulse for bandwidths of $\frac{2}{t_p}$ and $\frac{8}{t_p}$ respectively; (b) therefore corresponds to the inclusion of all the sidebands of Fig. 660 as far as the first zero on each side of the carrier; in (c) the sidebands between the fourth zeros are included. The leading edge of the reproduced pulse is curved and the convention of Fig. 663 is adopted for estimating the delay time and time of rise. If the input voltage v_i were sinusoidal, of frequency f_0 , the output would have amplitude $\hat{v}_0 = |m| \hat{v}_i$.

This amplitude is used as

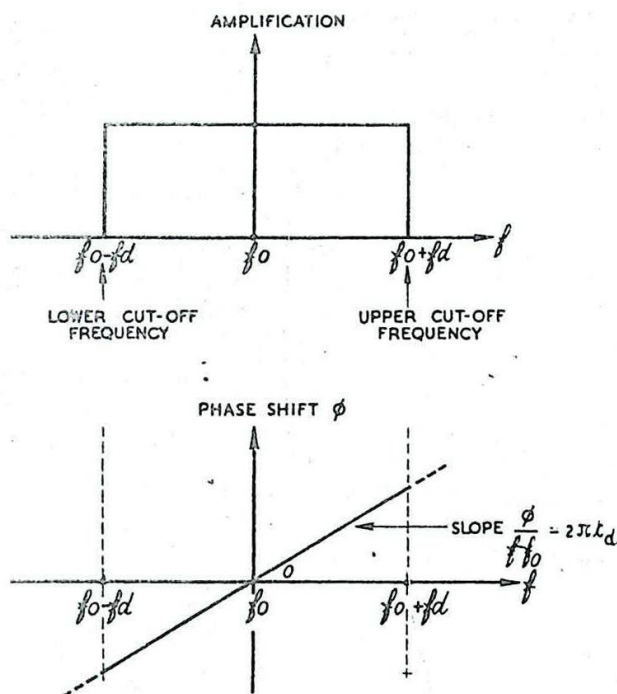


Fig. 661 - Amplitude and Phase-Shift characteristics of idealised band-pass amplifier.

a reference level when the input voltage is in the form of a RF pulse, also of amplitude \hat{v}_i , and the point X is taken when the pulse amplitude has risen to $\frac{1}{2} |\hat{m}| \hat{v}_i$. The tangent

at this point intersects the zero and $|\hat{m}| \hat{v}_i$ levels at P and Q respectively. The time difference between these two points is taken to be the time of rise t_r of the reproduced pulse, and the delay time t_d is taken as the time difference between X and the leading edge of the input pulse.

With these definitions it may be shown that :-

(1) $t_d = \frac{1}{2\pi} \times$ (slope of phase characteristic of Fig. 661).

(2) $t_r = \frac{1}{B}$, where B is the bandwidth, $2f_d$ (Fig. 661).

(3) the oscillations at the top of the pulse of Fig. 662 (c) have a frequency of $\frac{B}{2}$ c/s.

The response of the idealised amplifier discussed above is in the main close to that of a practical amplifier, i.e., the delay, time of rise and frequency of oscillations are much the same in the two cases. However, in practice, each successive pulse may be treated as distinct from those preceding it, since the pulses are not coherent, and if this is done it is obvious that no oscillations due to this pulse can occur before $t = 0$, i.e., there can be no output before the input is applied.

Investigation into the transmission properties of practical amplifiers shows that, if the amplification characteristic displays sharp cut-offs, the phase-shift characteristic is excessively non-linear at the extremities of the pass band. As a consequence, pronounced oscillations (overshoots) occur during and after the time interval of the reproduced RF pulse; (Fig. 664(a)). In radar systems, the overshoots, in so far as they constitute deviations from a true rectangular shape, should be

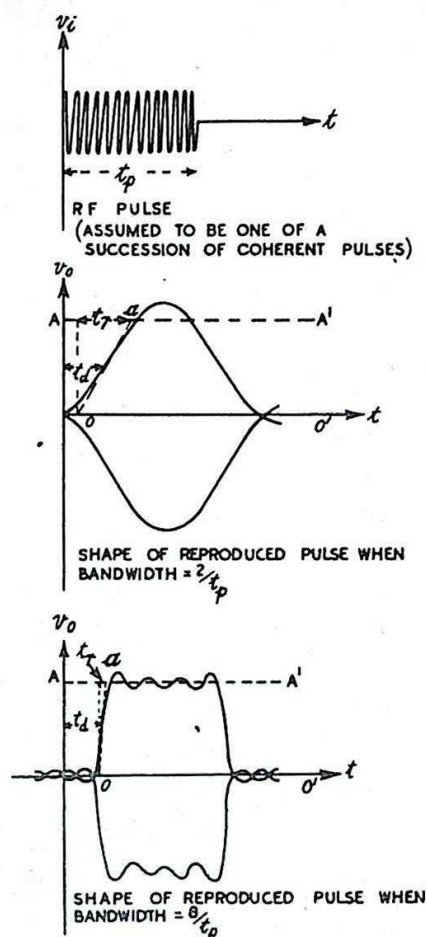


Fig. 662 - Distortion of a RF pulse by an idealised band-pass amplifier.

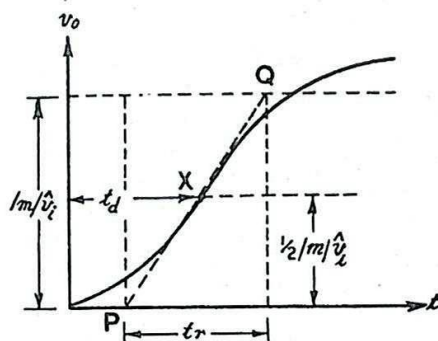


Fig. 663 - Delay time and time of rise of reproduced pulse.

avoided. This is particularly important in systems in which the target is indicated by a change in the brilliance of the trace on the CRT. The effect of such overshoots on the display can be minimised by the processes of Pulse Limitation and Clamping (see Chaps. 9 and 12). If the amplification characteristic of the amplifier displays gradual cut-offs, the phase-shift characteristic is not excessively non-linear, and the overshoots are comparatively small in amplitude; Fig. 664(b). In general, however, for a given pass band, a more gradual cut-off in the amplification characteristic means a slower rise in the leading edge of the reproduced pulse.

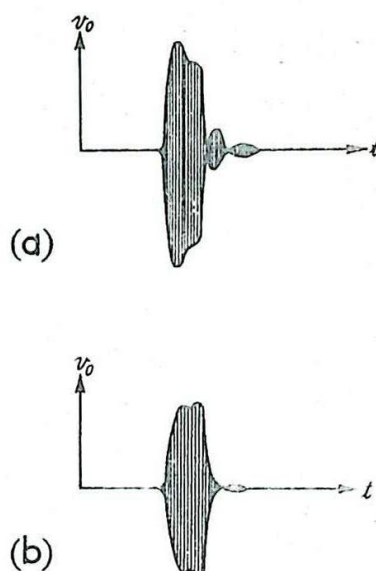


Fig. 664 - Reproduction of a rectangular RF pulse by a RF amplifier with (a) sharp cut-off and (b) gradual cut-off.

4. The Effect of Noise

All radar receivers possess inherent noise voltages which are in general of random amplitude and phase throughout all frequencies. Since the noise energy present in a given circuit is in most cases proportional to the bandwidth over which the circuit is responsive, the RMS value of the noise voltage is proportional to the square root of the bandwidth. The phase characteristic does not enter into the calculation of the average noise power since the phases of the noise voltages are purely random and therefore their effect, averaged out over an interval of time, is the same regardless of the phase-shift characteristic of the circuit. The noise voltages are amplified with the signal and produce on an A-type display the typical picture already described in Chap. 15 and illustrated in Fig. 642. In general, if the amplitude of a received pulse, applied to the deflecting plates of a CRT, is less than that of the noise voltages, it cannot be seen through the noise. This limits the maximum Range at which the equipment can detect signals. To increase the maximum Range, and to make the indication of signals more reliable at all ranges, it is clearly desirable to reduce as far as possible the noise voltages relative to the signal voltages.

Suppose that a RF pulse of duration t_p is applied to a circuit, the amplification of which may be kept constant as its bandwidth is varied. Starting with a narrow bandwidth the noise voltages will be small, but the rate of rise of the leading edge of the reproduced pulse will be so slow that the pulse will not have sufficient time to build up to its full amplitude (Fig. 665). As the bandwidth is increased the amplitudes of both pulse and noise increase also but the ratio of pulse to noise increases at first, passes through a maximum and then decreases; (Fig. 666). The optimum condition, $B = B_0$, depends to some extent on the type of display used, the recurrence frequency, etc.; for optimum detection of small signals on an A-display, B_0 is usually chosen between $\frac{1}{t_p}$ and $\frac{2}{t_p}$.

If it is desirable in a radar equipment to make accurate range measurements or to discriminate between closely neighbouring echoes,

faithful reproduction of the pulse shape is of greater importance than optimum signal/noise ratio, and the bandwidth used in such equipments may be as great as $\frac{10}{t_p}$. Naturally

this accuracy of pulse reproduction is obtained only by a sacrifice of maximum Range. A useful working compromise between fidelity of pulse shape and accuracy of Range measurement is to choose a bandwidth of about twice the optimum, i.e., an equipment transmitting a 1 microsecond pulse will have a receiver bandwidth of about 2-4 Mc/s.

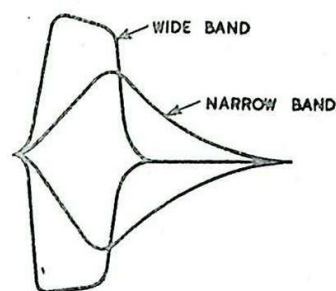
The choice of pulse duration will again depend on the purpose of the equipment. For any given pulse length t_p the bandwidth for optimum signal/noise ratio is proportional to $\frac{1}{t_p}$ so that a

relatively long pulse implies a narrow bandwidth with a corresponding reduction in noise. The use of a long pulse of the same peak power will therefore lead to an improvement in signal/noise ratio and an extension of maximum Range. In general, early warning equipments have fairly long pulses and narrow receiver bandwidths but poor discrimination, whereas for good Range discrimination the pulse is short, the receiver bandwidth wide and the Range limited.

5. The Superheterodyne Principle

It has been stated that radar receivers normally employ the superheterodyne principle. This is used in order to obtain the necessary high gain with stability. Further, on centimetre wavelengths no useful signal frequency amplification is possible, so that a change of frequency is essential. At the same time, selectivity is not a problem as it is in communications working, and the receiver bandwidth is determined solely by pulse fidelity and signal/noise considerations and not by the necessity of avoiding interference from transmissions on adjacent channels. Similarly, the choice of the intermediate frequency is not dictated by second-channel interference problems. This is not to say that radar receivers are immune from interference, but such interference is of a specialised character and is not removed by merely restricting receiver bandwidth.

The effect of the frequency-changer is to shift the pulse spectrum of Fig. 660 so that it is no longer centred about the signal frequency but about the intermediate frequency, the relative amplitudes and positions of the sidebands remaining unchanged. The considerations discussed above in connection with the bandwidth of circuits designed to pass RF pulses apply equally well to circuits used for amplification at the intermediate frequency. In fact it is in the IF amplifier that the necessary bandwidth is obtained; the signal frequency circuits, owing to



ENVELOPE OF REPRODUCED PULSE SHOWING EFFECT OF REDUCTION OF BAND WIDTH AT CONSTANT GAIN.

Fig. 665 - Envelope of reproduced pulse showing effect of reduction of bandwidth at constant gain.

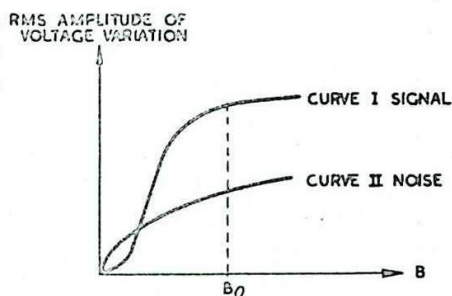


Fig. 666 - Variation of signal and noise output voltages with bandwidth of circuit.

the higher working frequency are usually much more flatly tuned. This corresponds with normal communications practice.

The necessity for amplifying a short RF pulse and the consequent wide bandwidth needed restricts the choice of the intermediate frequency. For a given bandwidth and tuning capacitance the gain of a RF amplifier is independent of the mid-band frequency so that stability considerations suggest the use of a low value of IF: on the other hand there must be a sufficiently large number of IF cycles to "fill in" the pulse envelope. At least 20 cycles are needed for this purpose, so that for the accurate reproduction of 1 microsecond pulses an IF of at least 20 Mc/s. would be needed. Longer pulses could be dealt with at correspondingly lower frequencies, e.g. a 5-microsecond pulse could be handled by an IF amplifier working on 4 Mc/s. Where accurate reproduction of the pulse shape is not required a smaller number of IF cycles can be tolerated, as few as 5 cycles per pulse being satisfactory in some cases.

A further advantage of not using a high IF is that certain sources of noise are thereby avoided; (see Chap. 15 Sec. 10).

The output from the IF amplifier is applied to the detector, the output of which should ideally consist of a rectified pulse having the same shape as the envelope of the IF pulse. Since all sideband components up to $\frac{B}{2}$ on each side of the carrier have been retained in the

IF amplifier the output from the detector may be analysed into a frequency spectrum extending from zero up to an upper limit of $\frac{B}{2}$ c/s. For distortionless detection these components should be maintained with correct relative amplitudes and correct relative phase. However, the detector must always incorporate some filter device to remove the residual intermediate frequency components, together with harmonics which are inevitably produced in the detection process. The higher the IF chosen the easier it is to filter out the unwanted IF components and their harmonics without introducing amplitude or phase changes in the frequency range $0 - \frac{B}{2}$ c/s. The choice of at least 20 IF cycles per pulse enables this to be done without difficulty.

Similarly the video-frequency stages following the detector should reproduce the output of the latter without distortion; i.e. the video stages should have uniform amplitude/frequency and linear phase/frequency characteristics up to at least $\frac{B}{2}$ c/s.

6. Typical 200 Mc/s Radar Receiver

The following are the details of a typical radar equipment:-

- (1) Signal frequency : 200 Mc/s.
- (2) Intermediate frequency : 45 Mc/s.
- (3) Pulse length : 3 microseconds. (There are thus 135 IF cycles per pulse.)
- (4) IF bandwidth $B = 3$ Mc/s. (With this value of $\frac{10}{t_p}$, accurate reproduction of the pulse is obtained, the time of rise of the leading edge being 0.33 microseconds.)
- (5) Video amplifiers : Amplitude/frequency characteristic flat up to 2 Mc/s. (Since only sidebands up to 1.5 Mc/s. on each side of the carrier are passed by the IF amplifier the video amplifier need only handle frequencies up to 1.5 Mc/s., but the extension of the video frequency range beyond that actually required improves the amplifier characteristics. In particular, the phase characteristic tends to become excessively non-linear at the upper

end of the range; with 2 Mc/s. upper frequency limit this non-linear portion is outside the band of frequencies handled by the amplifier.)

INFLUENCE OF NOISE FACTORS ON RF AMPLIFICATION

7. General

In Sec. 4 some of the more general problems relating to noise in receivers are discussed, showing how unavoidable noise sets an upper limit to the maximum Range obtainable with a radar equipment. In this section we consider in greater detail how noise affects the design of RF amplifiers and of valves used in such amplifiers. The term Radio Frequency will be taken to include both Signal Frequency and Intermediate Frequency. The effects of noise in Frequency Changers will be dealt with in Secs. 24-28.

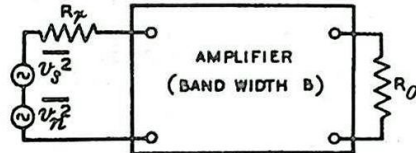


Fig. 667 - Block diagram of ideal RF amplifier.

8. The Ideal RF Amplifier

In an ideal RF amplifier all the valves and circuits would be completely free of noise and the signal/noise ratio at the output of such an amplifier would depend on the aerial alone. Consider the block diagram of Fig. 667, the amplifier being considered to be free from noise and to have ideal amplitude/frequency and phase/frequency characteristics over an adequate band-width B. The output impedance of the aerial, its radiation resistance R_r (see Chap. 17 Sec. 11), and the input and output impedances of the amplifier, R_i and R_o , are assumed to be pure resistances over the bandwidth B. The input circuit is shown in Fig. 668. During the pulse the signal power present in R_o will be $|m|^2$ times the signal power present in R_i where $|m|$ is the amplification. Similarly the noise power in R_o will be $|m|^2$ times the noise power in R_i associated with the bandwidth B since only those frequency components of noise lying within the pass band can affect the output.

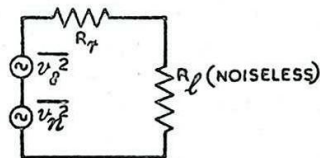


Fig. 668 - Input circuit of ideal RF amplifier.

Let $\overline{v_s^2}$ be the mean square signal voltage in the aerial during the pulse. The signal power P_s is then given by

$$P_s = \frac{R_i}{(R_r + R_i)^2} \cdot \overline{v_s^2} \dots\dots\dots (1).$$

Similarly, the noise power P_n , is given by

$$P_n = \frac{R_i}{(R_r + R_i)^2} \cdot \overline{v_n^2} \dots\dots\dots (2).$$

where $\overline{v_n^2}$ is the mean square noise voltage in R_r associated with the bandwidth B;

$$\text{i.e., } \overline{v_n^2} = 4kTB R_r = K R_r, \dots\dots\dots (3),$$

where $K = 4kTB$.

The ratio of noise/signal power at the input is thus

$$P_n/P_s = KR_r/v_s^2 \dots\dots\dots (4),$$

which is also the noise/signal power in R_o . This ratio (4) is independent of R_o so that the signal/noise ratio in an ideal amplifier is independent of the aerial matching conditions.

9. The Practical RF Amplifier: Noise Factor

In an actual amplifier R_o is noisy and there is also noise in the valves and other components, the net result being that the ratio of noise/signal power in the output is higher for a real amplifier than for an ideal one. This deterioration is expressed by the Noise Factor F which is defined as :-

$$F = \frac{\text{Noise/signal power in output of real amplifier} \dots\dots (5),}{\text{Noise/signal power in output of ideal amplifier}}$$

F , being in effect a power ratio, may be expressed either as a pure number or in decibels; e.g. $F = 4$ or $F = 6$ db.

In order that the noise factor of an amplifier may be as low as possible careful consideration must be given to the design of the first stage. Since the signal and noise outputs of this stage in particular are amplified together by all succeeding stages it is clearly desirable that the noise/signal ratio at the first grid should be as low as possible.

Consider the circuit of Fig. 669 in which an aerial of resistance R_r is shown inductively coupled to the first tuned circuit of the amplifier by a perfect transformer, of turns ratio $1:n$. In practice the transformer will not be perfect, the circuit will not necessarily consist of lumped inductance and capacitance and some reactance may be present in the aerial. These differences will not, however, invalidate the argument which follows; they will merely complicate the mathematics without affecting the general conclusions.

10. Input Circuit Coupling with Noiseless First Valve

In communications practice where noise is not the limiting factor the turns ratio n is usually chosen to give maximum signal at the grid. This condition is obtained when the aerial resistance is matched to the dynamic resistance R_d of the tuned circuit, i.e. when

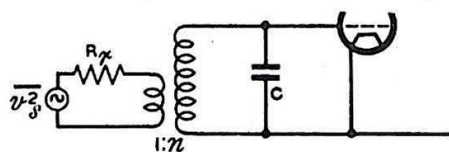


Fig. 669 - Input circuit of first stage of amplifier.

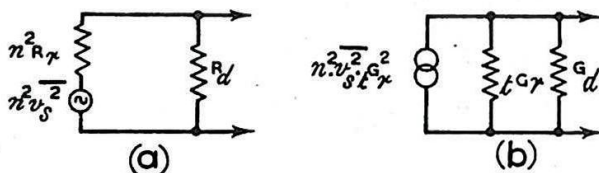


Fig. 670 - Equivalent signal circuits of fig. 669.

$$n^2 R_r = R_d \dots\dots\dots (6)$$

The equivalent circuits of Fig. 669 are shown in Fig. 670, the current circuit (b) being the easier to work with when noise calculations are required, since the use of conductances instead of resistances leads to a simplification of the algebra. t_{Gr} is the Transferred Aerial Conductance and is equal to $\frac{1}{n^2 R_r} = \frac{G_r}{n^2}$. From Fig. 670 (b) the mean square signal voltage at the grid is

$$\overline{v_i^2} = n^2 \cdot \overline{v_s^2} \cdot t_{Gr}^2 \cdot \frac{1}{(t_{Gr} + G_d)^2},$$

$$= \frac{G_r}{(t_{Gr} + G_d)^2} \cdot \frac{t_{Gr}}{\overline{v_s^2}} \dots \dots \dots (7)$$

and this is a maximum when $t_{Gr} = G_d$. This is the same result as that given by equation (6), in terms of conductances instead of resistances.

It is now possible to see how the coupling between aerial and tuned circuit needs to be modified when noise is present. Considering the noise in the tuned circuit alone, i.e. assuming for the moment that the valve is noiseless, the mean square noise voltage $\overline{v_i^2}$ at the grid within the band-width B is found from the equivalent circuit of Fig. 671. It is

$$\overline{v_i^2} = \frac{K (t_{Gr} + G_d)}{(t_{Gr} + G_d)^2} \dots \dots \dots (8)$$

Dividing (8) by (7), we obtain

$$\frac{\overline{v_i^2}}{\overline{v_s^2}} = \frac{K (t_{Gr} + G_d)}{t_{Gr} \cdot G_r \cdot \overline{v_s^2}} \dots \dots \dots (9)$$

but for the ideal amplifier

$$\frac{\overline{v_i^2}}{\overline{v_s^2}} = \frac{K R_r}{\overline{v_s^2}} = \frac{K}{G_r \overline{v_s^2}} \dots \dots \dots (10)$$

from (4).

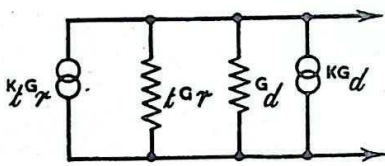


Fig. 671 - Equivalent noise circuit of fig. 669 (valve assumed noiseless).

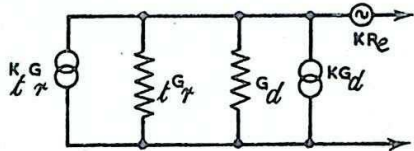


Fig. 672 - Equivalent noise circuit of fig. 669 with noisy valve.

By definition, the noise factor F is the quotient of (10) and (9) so that

$$F = \frac{t_{Gr} + G_d}{t_{Gr}} = 1 + \frac{G_d}{t_{Gr}} \dots \dots \dots (11)$$

It follows from (11) that there is no optimum coupling from the noise point of view, the value of F decreasing asymptotically to unity as t_{Gr} increases. For the matched condition, in which maximum signal is obtained at the grid, the noise factor F is equal to 2. This condition should now be avoided and the transferred aerial conductance made as large as possible. The signal naturally suffers but the all-important ratio of signal-to-noise is improved.

In RF amplifiers used in radar receivers it is customary to connect resistors in parallel with the tuned circuits to obtain the necessary band-width. In the case of the first tuned circuit this should not be done, for such a resistor would increase the dynamic conductance G_d of the tuned circuit and thus increase the noise factor; (equation (11)). The damping imposed by the transferred aerial conductance is usually sufficient to give an adequate band-width.

11. Input Circuit Coupling with Noisy First Valve

As an introduction to the effects of valve noise it will now be assumed that shot and partition noise are present in the first stage. The circuit of Fig. 671 is then modified by the inclusion of the valve noise generator between the circuit and the grid; (Fig. 672). Since this generator is a purely fictitious one, i.e. there is no resistance of value R_e anywhere in the circuit, it cannot be transformed into an equivalent current generator and similarly there is no advantage in expressing the equivalent noise resistance as a conductance. The mixture of current and voltage generators and conductances and resistances of Fig. 672 not only simplifies the algebra but also emphasizes the distinction between resistances which represent energy losses and those which are merely convenient mathematical devices for representing noise magnitudes.

It follows from Fig. 671 and 672 that the mean square noise voltage at the grid in the latter case is

$$\overline{v_i^2} = \frac{K (t_{Gr} + G_d)}{(t_{Gr} + G_d)^2} + K R_e \dots\dots\dots (12)$$

The signal voltage is unchanged by the introduction of the additional noise generator so that the mean square signal voltage is

$$\overline{v_s^2} = \frac{G_r t_{Gr}}{(t_{Gr} + G_d)^2} \cdot \overline{v_s^2} \dots\dots\dots (13)$$

as before.

Dividing (12) by (13) and then by (10) we obtain the noise factor :-

$$F = 1 + \frac{G_d}{t_{Gr}} + \frac{(t_{Gr} + G_d)^2}{t_{Gr}} \cdot R_e \dots\dots\dots (14)$$

It is now found that there is an optimum value of G_r , giving a minimum noise factor. If (14) is differentiated and the derivative equated to zero the optimum value of t_{Gr} is found to be :-

$$t_{Gr} (\text{opt}) = G_d \sqrt{\frac{1+x}{x}}; \text{ where } x = G_d R_e \dots\dots\dots (15)$$

and the minimum noise factor is :-

$$F = 1 + 2x + 2\sqrt{x + x^2} \dots\dots\dots (16)$$

It follows from (15) that the best value of t_{Gr} is always greater than that required for matching ($t_{Gr} = G_d$) since $\frac{1+x}{x}$ is always greater than unity.

The noisier the valve, the greater the value of x and the smaller the optimum t_{Gr} . In the extreme case where the valve noise over-rides the circuit noise, (15) reduces to $t_{Gr} = G_d$, which expresses the obvious conclusion that when valve noise predominates the best signal/noise ratio is obtained when the signal at the grid is a maximum.

With a given value of equivalent noise resistance R_e the variation of noise factor and band-width with t_{Gr} is shown in Fig. 673. Curve F(I) shows the change of noise factor with t_{Gr} when valve noise is absent, while curve F(II) shows the corresponding case in which $R_e = 1400$ ohms, a value which is appropriate to a CV1091 valve. It will be noted that a transferred aerial conductance which is slightly greater than the optimum is needed in the latter case to give a band-width of 4 Mc/s, the deterioration in noise factor due to this being only 0.15 db. When other sources of damping which have so far been neglected are taken into account the band-width is more than adequate when the noise factor is a minimum.

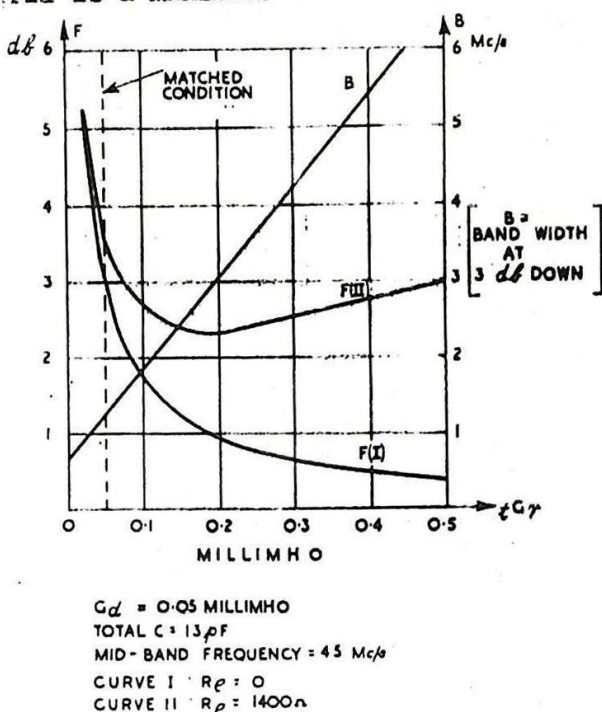


Fig. 673 - Noise factor and band-width of circuit of fig. 672.

We may use the curves of Fig. 673 to compare the noise factors under matched and optimum conditions. For the matched aerial $F = 3.62$ db, whereas the minimum noise factor is 2.56 db. This improvement of 1.26 db is equivalent to a 34% increase in transmitter power.

12. Effects of Transit Time and Cathode Lead Inductance

It has been shown in Chap. 7 Sec. 25 that at high frequencies, when the time of transit of an electron through the valve becomes an appreciable fraction of the input cycle, energy is absorbed from the input circuit. Alternatively it may be said that the valve has a finite input conductance due to transit time. This transit time input conductance G_t is effectively in parallel with the tuned circuit, causing both a loss of signal and an increase in bandwidth. As stated in Chap. 15 Sec. 10, the mean square noise current at the grid within the frequency band B due to transit time only (Induced Grid Noise) may usually be taken as

$$\overline{i_{gn}^2} = 4kTB \cdot 4 \cdot 8G_t \quad \dots \dots \dots (17)$$

so that the transit time input conductance is about five times as noisy as an equal physical conductance at the same temperature. The effect of transit time in the valve on the circuit of Fig. 669 is to add the noisy

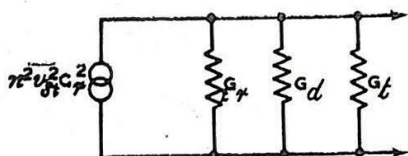


Fig. 674 - Equivalent signal circuit of fig. 670(b) with addition of transit time conductance.

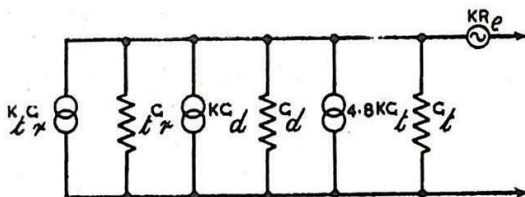


Fig. 675 - Equivalent noise circuit of fig. 672 with addition of transit time (induced grid) noise.

transit time conductance in parallel with the tuned circuit, giving the equivalent signal and noise circuits of Figs. 674 and 675 respectively.

When transit time noise and shot noise are combined as in Fig. 675 it is not at once clear that the linear addition of mean square voltages is valid, since such addition is strictly true only for random independent events such as noise contributions from two distinct sources. Now transit time and shot noise are not independent, both being produced by the same electron stream; however, a more detailed study of the frequency spectrum of the noise shows that for any given component of this spectrum the transit time and shot noise are effectively in phase quadrature, at least for the small transit angles normally encountered in practice. The mean square voltages may thus be added linearly to give the total effect, just as if they were produced by independent sources.

In any valve operated with its cathode nominally earthed (grounded-cathode circuit) it is essential to have a certain length of lead between cathode and earth, if only to prevent undue conduction of heat from the hot cathode to the seal through the glass. This cathode lead possesses inductance which affects the input impedance of the stage.

Referring to Fig. 676, if L_K is the inductance of the cathode lead and C_{gk} the grid-cathode capacitance, then it has been shown in Chap. 7 Sec. 25 that the input conductance is :-

$$G_K = \omega^2 \cdot L_K \cdot C_{gk} \cdot G_m$$

..... (18).

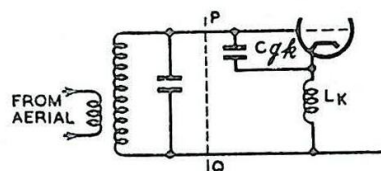


Fig. 676 - Input circuit of grounded-cathode stage showing cathode lead inductance.

This is the conductance which is in parallel with the input circuit at the dotted line PQ of Fig. 676. The input circuit to the left of PQ may be replaced by a current generator in parallel with the output conductance G_0 so that the equivalent signal circuit is that of Fig. 677.

It is now necessary to consider what noise generator (if any) should be associated with conductance G_K . A distinction must be made between the shot noise and partition noise which have hitherto been lumped together in a single noise generator, since the shot noise currents are present in the cathode lead whereas the partition noise currents are not. (See Chap. 15 Sec. 15). If the frequency spectrum of the shot noise is considered, any one component of this spectrum will produce a voltage across L_K (Fig. 676), and will thus be injected into the series circuit formed by L_K , C_{gk} and the input circuit. A consideration of the amplitude of the resultant current shows that the voltage developed across C_{gk} produces negative feedback of the noise and that the magnitude of this feedback is correct if the shot noise generator is placed as in Fig. 678 and the input conductance G_K is noiseless. The noise present in the circuit to the left of PQ in Fig. 676 is represented by the current generator i_{iN}^2 and R_p , the partition noise generator, is placed in its usual position, since such noise is not subject to feedback.

Suppose that for the moment partition noise is neglected; then the mean square signal voltage at the grid (Fig. 677) is:

$$\overline{e_v^2} = \overline{i_{iS}^2} \cdot \frac{1}{(G_p + G_K)^2} \quad \text{..... (19),}$$

while the mean square noise voltage (Fig. 678) is:

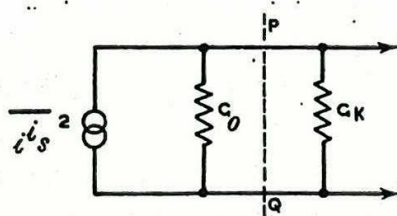


Fig. 677 - Equivalent signal circuit of fig. 676.

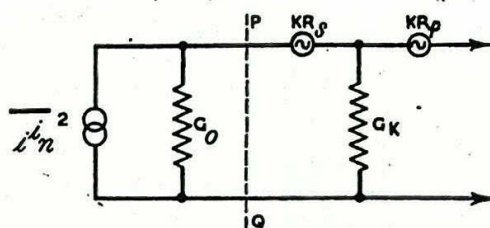


Fig. 678 - Equivalent noise circuit of fig. 676.

$$\overline{n^2} = \overline{i_n^2} \cdot \frac{1}{(G_0 + G_K)^2} + \frac{KR_s G_0^2}{(G_0 + G_K)^2} \dots\dots\dots(20).$$

The noise/signal ratio is therefore:

$$\frac{\overline{n^2}}{\overline{i_s^2}} = \frac{\overline{i_n^2}}{\overline{i_s^2}} + \frac{KR_s G_0^2}{\overline{i_s^2}} \dots\dots\dots(21),$$

which is independent of G_K and is therefore the same as if G_K were absent.

Since the current generator of Fig. 678 includes the effect of transit time it follows that the inverse feedback due to cathode lead inductance affects all sources of noise except partition noise.

It should be remembered that the conductance G_0 of Figs. 677 and 678 is made up of the transferred aerial conductance, the dynamic conductance of the tuned circuit and the transit time conductance; in fact the portions to the left of PQ are a condensed representation of Figs. 674 and 675 respectively. The complete circuits are thus as shown in Figs. 679 and 680.

From Fig. 679 the mean square signal voltage at the grid is :-

$$\overline{v^2} = \frac{n^2 \cdot v_s^2 \cdot t_{Gr}^2}{G^2}, \text{ where } G = t_{Gr} + G_d + G_t + G_K \dots\dots\dots(22),$$

and the mean square noise voltage (Fig.680) is

$$\overline{n^2} = \frac{K(t_{Gr} + G_d + 4 \cdot 8 G_t)}{G^2} + K R_s \frac{(t_{Gr} + G_d + G_t)^2}{G^2} + K R_p \dots\dots\dots(23).$$

The noise factor is given by :-

$$F = 1 + \frac{G_d}{t_{Gr}} + 4 \cdot 8 \frac{G_t}{t_{Gr}} + \frac{(t_{Gr} + G_d + G_t)^2}{t_{Gr}} \cdot R_s + \frac{G^2}{t_{Gr}} \cdot R_p \dots\dots\dots(24).$$

Equation (24) shows that the conductance G_K affects the noise factor only by its presence in the total conductance G in the last term, and this term is the one involving the equivalent partition noise resistor R_p .

The curves of Fig. 681 have been calculated for a CV1091 valve at a frequency of 45 Mc/s, the appropriate values being :-

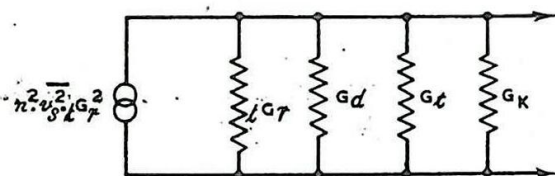


Fig. 679 - Equivalent signal circuit of fig. 674 with addition of conductance due to cathode lead inductance.

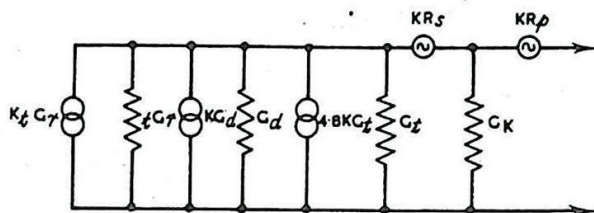


Fig. 680 - Equivalent noise circuit of fig. 675 with addition of noiseless conductance due to cathode lead inductance.

$$G_d = 0.05 \text{ millimho}$$

$$G_t = 0.045 \text{ "}$$

$$G_K = 0.065 \text{ "}$$

$$R_s = 400 \text{ ohms}$$

$$R_p = 1100 \text{ ohms}$$

$$\text{Total } C = 13 \text{ pF.}$$

Owing to the increased damping due to G_t and G_K the matched condition is now given by :-

$$t G_r = G_d + G_t + G_K \dots\dots\dots (25),$$

i.e. $t G_r = 0.16$ millimho, and the noise factor for this power match is 5.14 db. The minimum noise factor is 4.28 db and this is obtained for $t G_r = 0.5$ millimho. The band-width at the minimum noise condition is 8.4 Mc/s and this should be compared with the simple case of Fig. 673. The wider band is mainly due to the increased $t G_r$ needed to give minimum noise factor, but is also partly due to the greater damping caused by the finite input conductance of the valve.

13. Operation at Higher Frequencies

The input conductances due to transit time and the inductance of the cathode lead both vary as the square of the frequency so that if this frequency is increased from 45 Mc/s to, say, 180 Mc/s,

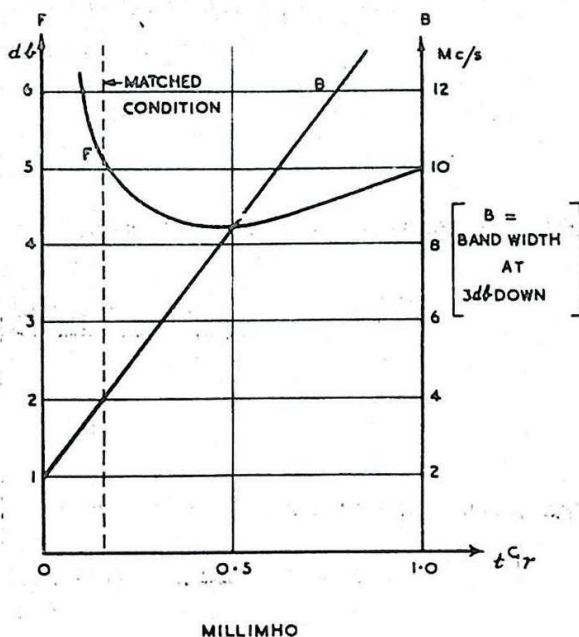


Fig. 681 - Noise factor and bandwidth of circuit of fig. 669 including effects of transit time and cathode lead inductance.

both conductances are increased by a factor of 16. G_t at this new frequency thus becomes 0.72 millimho and G_K 1.04 millimho. Assuming that it is possible by suitable circuit construction to keep the dynamic conductance of the tuned circuit at about the same value although the frequency is increased, this dynamic conductance will be so much less than the input conductance of the valve at 180 Mc/s. that it will have little effect on the noise factor and bandwidth.

If (24) is differentiated and the result equated to zero the transferred aerial conductance for minimum noise factor is found to be :-

$$t_{Gr}^2 = \frac{(G_t + G_d)^2 R_s + (G_t + G_d + G_K)^2 R_p + 4.8 G_t + G_d}{R_s + R_p}$$

..... (26).

Neglecting G_d and using the values for G_t and G_K previously quoted, viz: 0.72 and 1.04 millimhos respectively, the optimum value of t_{Gr} is 2.14 millimhos and the substitution of this in (24) yields the value 10.8 db for the minimum noise factor. This is about 2 db better than the experimental figure for the CV1091 at 200 Mc/s.

It is possible to calculate the contribution of partition noise to the noise factor. If it were practicable to eliminate partition noise, R_p would be zero in (24) and (26) and the optimum values of t_{Gr} and F would then become 3.1 millimhos and 6.0 db respectively. Partition noise is thus responsible for a 5.8 db loss in signal/noise, i.e. a virtual waste of 74% of the transmitter power. This fact leads to two possible lines of development: (i) the redesign of the pentode to reduce the partition noise as much as possible; (ii) the use of a suitable triode as a RF amplifier instead of a pentode.

Considering the first alternative: the high noise factor of the pentode is due to two interdependent causes: (1) the presence of partition noise; (2) the negative feedback due to inductance of the cathode lead which favours partition noise. In the CV1136, the CV1091 type of construction has been retained but the cathode is taken to four pins in the base instead of one. Each of these multiple cathode leads has about the same inductance as the single lead of the CV1091 so that the effective cathode lead inductance is only about one quarter. At the same time the screen current is reduced by aligning the grid and screen wires, thus reducing the partition noise; (see Chap. 15 Sec. 9). The equivalent partition noise resistance is thereby reduced to about 500 ohms instead of 1100 ohms for the CV1091. In practice, however, the CV1136 does not give the improvement in noise factor to be expected theoretically. Measurements have shown that the input conductances of the two valves are not substantially different, and the improvement, which is about 1 db at 45 Mc/s and 2 db at 180 Mc/s, is largely due to the reduction in partition noise. These are average values, and owing to production tolerances it is not unusual to find a particular CV1091 which gives a lower noise factor than a particular CV1136.

At frequencies of the order of 200 Mc/s the inductances of the screen and grid leads become important; the inductance of the latter, for example, is approaching series resonance with the grid-cathode capacitance so that the input impedance of the valve as seen from the pins is lower than that actually present at the electrodes. The simple circuit of Fig. 669 thus needs to be replaced by a fairly complicated network and the noise factor is no longer amenable to easy calculation. This accounts for the discrepancy of about 2 db between the calculated and measured noise factors of the CV1091 at 180 - 200 Mc/s and also, possibly, for the relatively small difference between the CV1091 and CV1136.

The presence of the screen lead inductance means that the screen is losing its effectiveness since its potential may no longer be regarded as constant, and in general the pentode ceases to be a useful amplifier above about 200 Mc/s.

14. The Grounded-Grid Triode

It has already been noted that the elimination of partition noise would bring about a considerable improvement in noise factor. A triode has no partition noise but cannot be used in the conventional grounded-cathode circuit owing to the feedback which would occur through the grid-anode capacitance. If, however, the triode is used in an inverted amplifier or grounded-grid circuit, the grid will, if a suitable valve construction is adopted, act as an electrostatic screen between the input and output circuits (Fig. 682).

It has been shown in Chap. 7 Sec. 20 that the input admittance of such a stage is

$$y_i = \frac{\mu + 1}{R_a + Z_l}, \dots\dots\dots (26),$$

where

μ is the amplification factor

R_a is the anode slope resistance

and Z_l the load in the anode circuit.

Over the pass band of the amplifier the anode load may be regarded as a pure resistance R_l so that the input admittance is thus a pure conductance of value

$$G_i = \frac{\mu + 1}{R_a + R_l} \dots\dots (27).$$

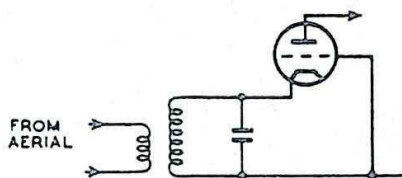


Fig. 682 - Input circuit of grounded-grid stage.

The signal circuit is therefore as shown in Fig. 683 and is the same as Fig. 679 with G_i replacing G_K . A

typical value of G_i is 5 millimhos and the effect of this in parallel with the tuned circuit is to produce such heavy damping that bandwidth is no longer a problem.

The shot noise currents flow through the input circuit and thus produce negative feedback voltages across it. These voltages are developed across a resistance so that the mechanism is different from that of the inverse feedback produced by cathode lead inductance in the circuit of Fig. 676. The result is the same, however, and it is found that the equivalent circuit of Fig. 680 is applicable to the grounded grid triode if G_K is replaced by G_i and the partition noise generator omitted (Fig. 684).

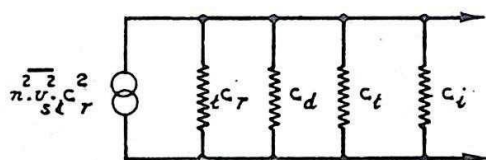


Fig. 683 - Equivalent signal circuit of fig. 682.

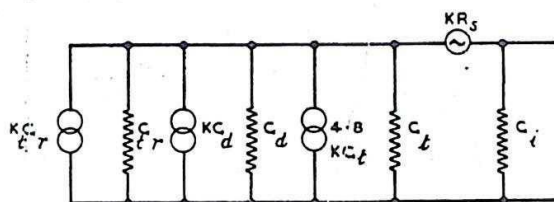


Fig. 684 - Equivalent noise circuit of fig. 682.

The noise factor of the stage is obtained from (24) by putting $R_p = 0$, and the optimum value of t_{Gr} is likewise obtained from (25). This has already been done in the discussion of the contribution of partition noise to the noise factor of a CV1091 at 180 Mc/s. The results were: $t_{Gr} = 3.1$ millimhos and $F = 6$ db. Thus if the transit time conductance is the same in the two cases (a not unreasonable assumption) a grounded grid triode gives the same noise factor as a pentode without partition noise. The replacement of a pentode by a grounded grid triode at 180 Mc/s would thus lead to an improvement of 5.8 db in signal/noise ratio, in other words, the virtual loss of 74% of the transmitter power would be eliminated.

In all the preceding discussions on noise factor the noise generated in stages other than the first has been neglected. Just as it is possible to represent the shot and partition noise of the first stage by an equivalent noise generator at the grid, even though these are produced in the anode circuit, so it is equally possible to represent all noise generated within the amplifier band-width by subsequent stages by a fictitious noise generator at the first grid. It is merely necessary to assign to this generator a mean square voltage which will produce the same noise power in the output load. The second stage of the amplifier will be the most important so that the mean square voltage of the equivalent noise source at the first grid is defined as KR_2 . This generator will be in series with the partition noise generator in Fig. 680 and will replace it in the corresponding equivalent circuit for the grounded grid triode. Fig. 684 is thus modified to Fig. 685 by the addition of this extra noise source.

The exact value to be assigned to R_2 will of course depend on the gain of the first stage as well as the noise generated by the second and following stages. A numerical example will illustrate this and emphasise the difference between a pentode and a grounded grid triode as the first stage. At 45 Mc/s a typical value for R_2 is 10 ohms; in the case of the CV1091 this is in series with the 1100 ohms partition

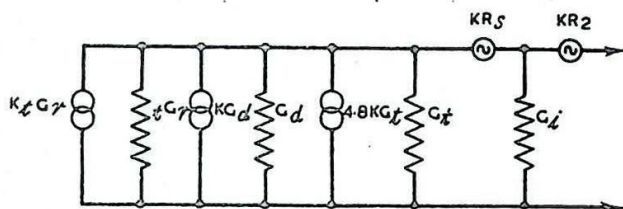


Fig. 685 - Modification of fig. 684 by the addition of equivalent noise generator for subsequent stages.

noise generator and is thus merely equivalent to increasing the latter by about 1%, an entirely negligible amount. On the other hand, in the grounded grid triode the signal and noise generated in the first stage are so reduced by the high degree of negative feedback that noise generated in the second stage begins to become important.

Re-writing equations (24) and (26) for the grounded-grid triode we obtain :-

$$F = 1 + \frac{G_d}{t_{Gr}} + 4.8 \frac{G_t}{t_{Gr}} + \frac{(t_{Gr} + G_d + G_t)^2}{t_{Gr}} \cdot R_s + \frac{G^2}{t_{Gr}} \cdot R_2$$

..... (28),

where $G = t_{Gr} + G_d + G_t + G_i$,

$$\text{and } t_{Gr}^2 = \frac{(G_t + G_d)^2 R_s + (G_t + G_d + G_i)^2 R_2 + 4.8 G_t + G_d}{R_s + R_2} \dots (29).$$

The low value of R_2 is to a large extent offset by the high value of G_i occurring in its coefficient. The variation of F with t_{Gr} is plotted in Fig. 686 for the two cases $R_2 = 0$ and $R_2 = 10$ ohms, the other values being:-

$$G_d = 0.05 \text{ millimho}$$

$$G_t = 0.05 \text{ millimho}$$

$$G_i = 5 \text{ millimhos.}$$

These are typical of the CV66 which is similar to the CV1091 and CV1136 in external appearance.

Referring to Fig. 686, owing to the high input conductance at the cathode the value of t_{Gr} for the matched condition is 5.1 millimhos and it is essential to have a considerably smaller value than this for minimum noise factor. This should be compared with Fig. 681 for the CV1091.

The improvement in noise factor between the CV66 and CV1091 is almost exactly 1 db; without the noise produced by the second stage the improvement would be 0.75 db greater.

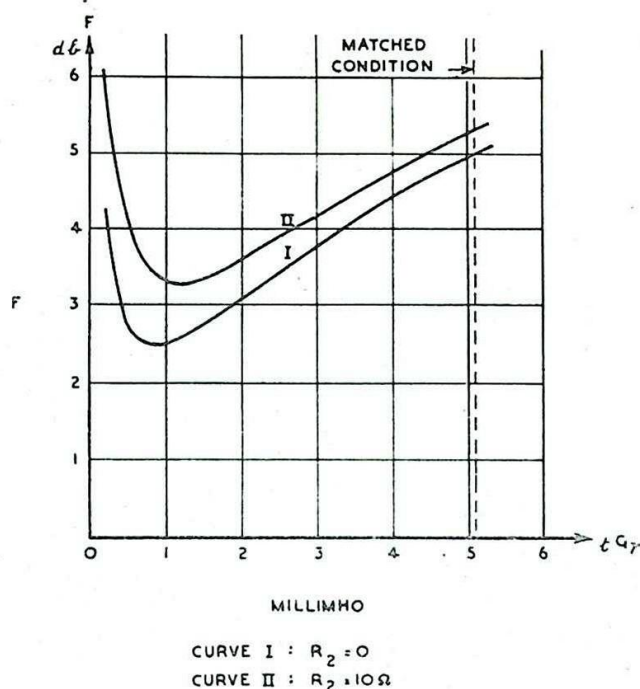


Fig. 686 - Variation of noise factor of circuit of fig. 682 with transferred aerial conductance.

As the frequency of operation is increased all stages of the amplifier become more noisy, so that not only does G_t increase but R_2 also. It is therefore not to be expected that a grounded grid triode will ever give a performance comparable with the theoretical pentode without partition noise. At 180 Mc/s the CV66 has a noise factor of about 7 db compared with 10.8 db and 6 db calculated for the CV1091 with and without partition noise respectively. At frequencies of the order of 600 Mc/s it is no longer possible to think of valve and circuit as separate entities; they must be designed together as a composite whole and every effort must be made to keep the electrode clearances to an absolute minimum in order to prevent transit time noise rising to an inordinate value. It should be remembered that the transit time conductance for any given valve increases by nearly 200 times between 45 Mc/s and 600 Mc/s. Valves such as the CV88 and CV153 which are grounded-grid triodes specially designed for high frequency working give noise factors of about 10 db at 600 Mc/s. Typical curves illustrating the variation in minimum noise factor with operating frequency for four valves are shown in Fig. 687. The circuit arrangements for different frequencies and the noise factor of the receiver as a whole are further discussed in Sec. 31.

FREQUENCY CONVERSION (MIXING)

15. General

Some reasons for employing the super-heterodyne principle in radar receivers are given in Sec. 5. These may be restated and amplified as follows :-

- (i) When centimetre waves are employed no useful signal frequency amplification is possible so that there is no alternative to the super-heterodyne principle, and the mixer must of necessity be the first stage in the receiver.
- (ii) It is much easier to obtain the necessary pass band characteristics if the amplification takes place in a fixed frequency or pre-tuned amplifier. This is normal communications technique but there is an even greater necessity for maintaining constant amplifier performance in radar reception (see Sec. 37).
- (iii) The problem of selectivity, so important in communications work, is considerably simplified by conversion of the signal to a lower "intermediate" frequency. However, in radar the question of distinguishing between two closely separated signal frequencies does not normally arise. Such interference as is experienced in radar working is usually of a different nature and is not dealt with in this chapter.

The higher the carrier frequency the greater the problem of obtaining amplification while maintaining stability. From this point of view it is advantageous to introduce the frequency converter at the earliest possible point in the receiver. However, the actual position occupied by the converter or "mixer" has to be considered from the point of view of obtaining the maximum signal/noise ratio in the receiver as a whole; (see Sec. 2).

It is intended to deal here with the basic principles of frequency conversion within the frequency range in which conventional valve structures (e.g. pentode, hexode) are used. Great detail will not

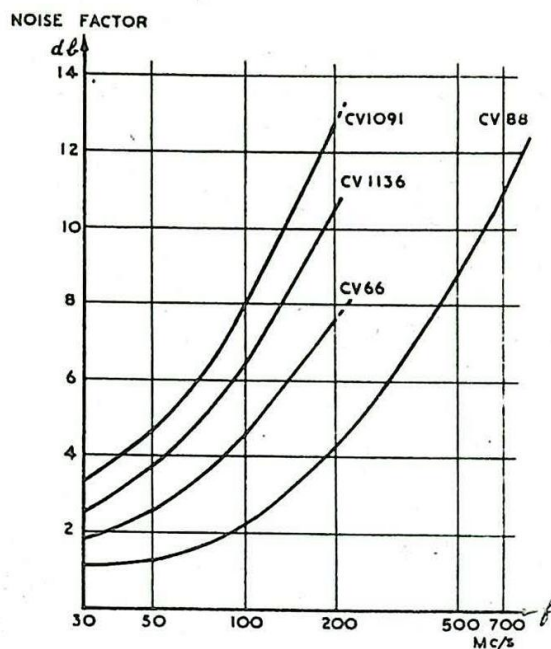


Fig. 687 - Variation of noise factor with operating frequency.

be entered into since such frequency converters have no extensive radar application and the subject is dealt with in many standard works. It does, however, form a useful introduction to the basic principles of the Two-Pole Converter (diode, crystal) which can be used at much higher frequencies. Neither pentode nor hexode mixers are commonly employed in Service radar receivers operating on a radio frequency higher than 50 Mc/s, since they are too noisy to use unless many stages of RF amplification have preceded the mixer stage. Triodes are usable as single-input mixers at frequencies up to about 200 Mc/s, but because of their large input capacitance they are replaced at higher frequencies by diodes. Diode mixers are commonly used for signal frequencies from 50 Mc/s - 500 Mc/s. The input capacitance and cathode lead inductance of diodes prohibits their use in the centimetre band; neither can be reduced indefinitely while sufficient emission is maintained, and ultimately crystal mixers must be employed. Although these are relatively inefficient frequency converters they are the only type which present sufficiently small input capacitance to the RF signal while still possessing a workable value of conversion efficiency.

16. Fundamental Principle of Frequency Conversion

The underlying principle of frequency conversion is to feed into the mixer (a) the incoming signal and (b) an output from an unmodulated Local Oscillator, i.e., an oscillator situated in the receiver. The output from the mixer contains a number of frequencies one of which is the required intermediate frequency.

In order to introduce into the output voltage or current frequencies which are not present in the input, the mixer must be a device which is effectively non-linear. For example, suppose that the signal voltage v_s applied to the mixer may be represented by :-

$$v_s = \hat{v}_s \cos \omega_s t \dots\dots\dots (1),$$

and that the local oscillator voltage v_o is given by

$$v_o = \hat{v}_o \cos \omega_o t \dots\dots\dots (2).$$

If these are added together the result is

$$v = v_s + v_o = \hat{v}_s \cos \omega_s t + \hat{v}_o \cos \omega_o t \dots\dots (3),$$

which is the voltage that would result if the signal and local oscillator output were applied in series to the mixer. Now let a resultant current i flow in the mixer output circuit as a result of the voltage v at the input; then if the mixer is linear,

$$i = b_1 v \text{ (where } b_1 \text{ is a constant)} \dots\dots\dots (4)$$

$$\text{or } i = b_1 \hat{v}_s \cos \omega_s t + b_1 \hat{v}_o \cos \omega_o t \dots\dots\dots (5),$$

showing that the output contains components of two frequencies only, namely, the signal and local oscillator frequencies, no new frequency having been produced.

The simplest non-linear case that may be considered is the addition of a term involving v^2 to (4), i.e.

$$i = b_1 v + b_2 v^2 \text{ (where } b_2 \text{ is a constant)} \dots\dots\dots (6).$$

Substituting from (3), we obtain

$$i = b_1 (\hat{v}_s \cos \omega_s t + \hat{v}_o \cos \omega_o t) + b_2 (\hat{v}_s \cos \omega_s t + \hat{v}_o \cos \omega_o t)^2$$

and, after expansion, this may be put in the form

$$\begin{aligned} i = & b_1 \hat{v}_s \cos \omega_s t + b_1 \hat{v}_o \cos \omega_o t \\ & + \frac{1}{2} b_2 (\hat{v}_s^2 + \hat{v}_o^2) \\ & + \frac{1}{2} b_2 \hat{v}_s^2 \cos 2\omega_s t + \frac{1}{2} b_2 \hat{v}_o^2 \cos 2\omega_o t \\ & + 2 b_2 \hat{v}_s \hat{v}_o \cos \omega_s t \cos \omega_o t \dots\dots\dots (7). \end{aligned}$$

The linear term in (6) gives signal and oscillator frequency components as before. The quadratic term gives rise to :-

- (1) a DC component;
- (2) second harmonic components of signal and oscillator frequencies;
- (3) a product term (the last in (7)) which on further expansion gives
 $b_2 \hat{v}_s \hat{v}_o \cos (\omega_s - \omega_o) t + b_2 \hat{v}_s \hat{v}_o \cos (\omega_s + \omega_o) t$

and therefore represents components of two new frequencies, which are respectively equal to the difference and sum of the signal and oscillator frequencies. (1) and (2) might have been expected, since a non-linear device is to some extent a rectifier, while operation on a curved characteristic always gives rise to harmonics. The terms involved in (3) are those fundamental to the frequency-changing process, the difference term usually being chosen for IF amplification. If equation (6) is extended further to include terms of higher order in v , viz :-

$$v = b_1 v + b_2 v^2 + b_3 v^3 + b_4 v^4 + \dots, \dots\dots\dots (8)$$

the output current will include not only signal and oscillator frequencies and their second harmonics but also a whole series of "mixing" components of angular frequencies $\pm (m\omega_s \pm n\omega_o)$ where m and n may take independently any integral values from zero upwards. If the signal is of very small amplitude compared with the local oscillator input (i.e., $\hat{v}_o \gg \hat{v}_s$) the only mixing components of appreciable amplitude are those for which $m = 1$, i.e. those of angular frequencies $\pm (\omega_s \pm n\omega_o)$. $n = 1$ gives the sum and difference frequencies which can be called "first order" mixing components, while the frequencies for which $n = 2, n = 3$, etc., may be called "second order", "third order", etc., respectively.

For a signal frequency f_s , local oscillator frequency f_o , and intermediate frequency f_i there is thus a general relation,

$$\begin{aligned} f_s \pm n f_o &= \pm f_i \\ \text{giving} \quad f_o &= \frac{f_s \pm f_i}{n} \dots\dots\dots (9). \end{aligned}$$

Thus for a signal frequency of 200 Mc/s and an IF of 45 Mc/s possible values for f_o are :-

- (a) 155 or 245 Mc/s (1st order mixing);
- (b) 77.5 or 122.5 Mc/s (2nd order mixing);
- (c) 51.7 or 81.7 Mc/s (3rd order mixing);

etc.

So far nothing has been said about the relative efficiencies of the different orders of mixing. This will depend upon the characteristic of the mixer, but in general it may be said that the efficiency falls off considerably with increasing order. (The efficiency can be roughly defined here as the IF output current for a given signal voltage input.) In practice therefore it is usual to employ first order mixing, although in some cases higher orders are used. For example, it may not be convenient to use an oscillator with a sufficiently high fundamental frequency to make first order mixing possible. Alternatively the LO frequency might be reduced in order to provide a wide tuning range, since the change in frequency of the n th harmonic is n times the change in the fundamental frequency. (A relatively wide tuning range is more easily obtained at lower frequencies, since the effects of stray capacitance, lead inductance, etc. are less important.)

Non-linear devices are readily available in the triode, pentode or hexode which have curved mutual characteristics. The analysis of frequency conversion with such devices is comparatively straightforward provided that it is assumed that the input impedance is infinite and that there is no coupling or interaction between input and output circuits. The inherent difficulties of the problem of the two-pole converter arise from the fact that in this case these assumptions are no longer valid.

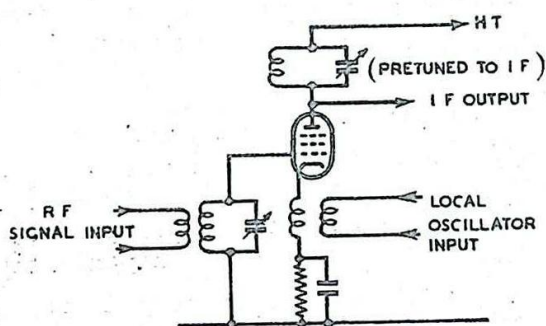


Fig. 688 - Pentode as single input mixer.

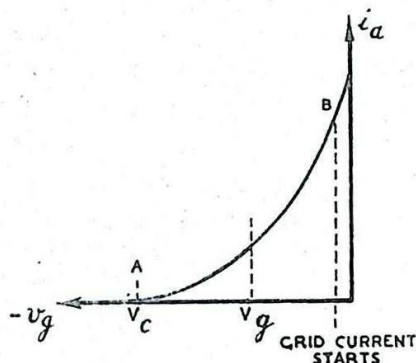


Fig. 689 - Mutual characteristic of pentode of fig. 688.

17. The Single-Input Mixer

A simple frequency converter (sometimes known as a Single-Input Mixer) using a pentode is shown in Fig. 688. The term Single Input is used to denote the application of both signal and oscillator voltages effectively in series between the cathode and one other electrode, in this case the control grid. The mutual characteristic is shown in Fig. 689 and it is assumed that the working point never runs the valve into grid current nor beyond cut-off, i.e. it always lies between A and B on the characteristic.

The analysis is simple if it is assumed that the characteristic between A and B is parabolic, so that the anode current for any given value of grid-cathode voltage V_g is given by :-

$$i_a = k_o (v_g - V_c)^2 \dots\dots\dots (10)$$

where V_c is the cut-off voltage. Bias of value V_g is applied to the valve so as to bring the mean position of the working point about half way between A and B. The instantaneous grid voltage is then

$$v_g = V_g + \hat{v}_s \cos \omega_s t + \hat{v}_o \cos \omega_o t \dots\dots\dots (11).$$

It is convenient to write

$$v_g = V_g + v \dots\dots\dots (12)$$

where $v = \hat{v}_s \cos \omega_s t + \hat{v}_o \cos \omega_o t \dots\dots\dots (13),$

and if $V_g + v$ is substituted for v_g in (10) the result may be written in the form :-

$$i_a = b_o + b_1 v + b_2 v^2 \dots\dots\dots (14).$$

Equation (13) is identical with (3) while (14) differs from (6) only by the presence of the constant term b_o , representing the direct current at the bias point. The resultant anode current is therefore obtained from (7) merely by adding b_o to the right hand side, and as before the anode current contains first order mixing components. Since the anode load presents an appreciable impedance at one frequency only, in this case the difference frequency, the voltage developed at the anode consists almost entirely of the component at this frequency.

A useful concept in mixers using valves of the conventional type is the Conversion Conductance G_c which is defined as :-

$$G_c = \frac{\text{peak value of IF current in anode circuit}}{\text{peak value of signal voltage applied to the valve}} \dots\dots\dots$$

This corresponds to the mutual conductance when the valve is used as a straight amplifier. Comparison of (10) and (14) shows that $b_2 = k$ while the amplitude of the IF component is $b_2 \hat{v}_s \hat{v}_o$ so that the conversion conductance is

$$G_c = k_o \hat{v}_o \dots\dots\dots (15).$$

Clearly the best conversion is obtained when the oscillator input voltage is the maximum possible which, if the working point is not to move beyond the limits AB, is, $V_g - V_c$, so that the maximum possible conversion conductance under these conditions is

$$G_c = k_o (V_g - V_c) \dots\dots\dots (16).$$

From (10) the mutual conductance G_m at the bias point is

$$G_m = 2k_o (V_g - V_c) \dots\dots\dots (17),$$

so that the process of conversion has only half the voltage efficiency of straight amplification. The figure of one-half naturally holds only for a parabolic characteristic under the stated operating conditions, but the general result is true for any single-input mixer, namely, that G_c is always less than G_m for the same valve and the same mean anode current.

In practice the mutual characteristic is not parabolic, also the working point may move beyond cut-off, so that the relation between i_a and v involves terms of higher degree than the second, viz :-

$$i_a = b_0 + b_1 v + b_2 v^2 + b_3 v^3 + b_4 v^4 + \dots \dots \dots (18).$$

Under these conditions higher order mixing components are present in the anode circuit.

The disadvantages of single-input mixers are well known. The oscillator voltage must be sufficient to swing the working point between the limits of the mutual characteristic otherwise conversion is inefficient. If the limits are exceeded grid current flows and damps the signal input circuit; hence the local oscillator input voltage is critical. Coupling may occur between the signal and oscillator circuits through the grid-cathode capacitance, causing interaction between the two tuning controls. The use of a multi-electrode valve with independent application of the signal and local oscillation to separate grids which are electrostatically screened from each other minimises this objectionable tendency as well as eliminating any chance of grid current damping the signal circuit. Such a system may be referred to as a Double-Input Mixer.

18. The Double-Input Mixer

An example of a double-input mixer is shown in Fig. 690. The signal is usually fed to the first grid and the local oscillation to the third; in some mixers these feeds are interchanged.

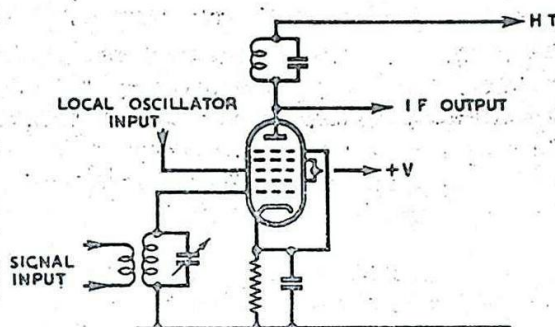


Fig. 690 - Typical double-input mixer circuit.

In considering the operation of this mixer it is essential to examine the implications of the term "effectively non-linear". The mixing action does not depend on curvature of the mutual characteristics connect-

ing anode current with either first grid voltage (v_1) or third grid voltage (v_3). These may be ideally represented by straight lines; (Fig. 691). If instead of the one characteristic of Fig. 691(a), a set is taken for different values of third grid voltage, the results are approximately as shown in Fig. 692. The characteristics are all straight lines with a common cut-off voltage but their slope depends on the voltage applied to the third grid. It is thus possible to write

$$i_a = G_1 (v_1 - 1V_c) \dots \dots \dots (19),$$

where G_1 , the mutual conductance referred to the first grid, is a parameter depending on v_3 , and $1V_c$ is the cut-off voltage. The simplest variation in G_1 which may be considered is linear, i.e.

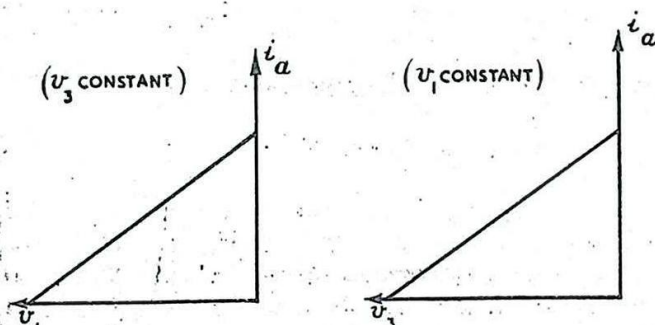


Fig. 691 - Ideal mutual characteristics of double-input mixer.

$$G_1 = A (v_3 - 3V_c) \dots (20)$$

which combined with (19) gives

$$i_a = A (v_1 - 1V_c) (v_3 - 3V_c) \dots \dots \dots (21).$$

Setting v_3 and v_1 constant in turn will give the linear relations expressed graphically in Figs. 691 (a) and (b). Both first and third grids carry bias in addition to their respective oscillatory inputs, so that :-

$$v_3 = V_3 + \hat{v}_0 \cos \omega_0 t \dots\dots\dots (22),$$

with a corresponding equation for v_1 .

Substitution in (21) and subsequent rearrangement yields the equation for the anode current in the form :-

$$i_a = a_1 + a_2 \hat{v}_s \cos \omega_s t + a_3 \hat{v}_0 \cos \omega_0 t + a_4 \hat{v}_0 \hat{v}_s \cos \omega_s t \cos \omega_0 t \dots\dots\dots (23).$$

The last term of this equation gives rise to the two first order mixing terms :-

$$\frac{1}{2} a_4 \hat{v}_s \hat{v}_0 \cos (\omega_s - \omega_0) t + \frac{1}{2} a_4 \hat{v}_s \hat{v}_0 \cos (\omega_s + \omega_0) t \dots\dots (24).$$

Thus, although both mutual characteristics are linear, frequency changing occurs.

There is no paradox in this. With a single input mixer there are two variables only, i_a and v_g , and the relation between them is represented by a curve drawn in two dimensions, i.e. the mutual characteristic. If this characteristic is straight, no frequency conversion occurs; if it is curved, frequency conversion takes place. In the case of a double-input mixer three variables are involved, i_a , v_1 and v_3 , and the relation between them can be expressed by a three-dimensional graph, i.e. a surface. The equation of this surface is that of (21), and although lines drawn on it for constant values of v_1 and v_3 are straight the surface itself is curved. If the surface were plane, equation (21) would be replaced by one of the form:-

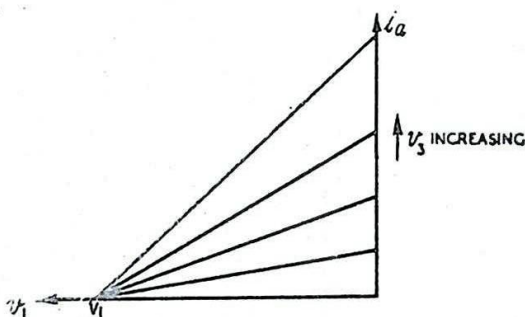


Fig. 692 - Variation of $i_a - v_1$ characteristic with v_3 .

$$i_a = p + q v_1 + r v_3 \dots\dots\dots (25),$$

which would give no frequency-changing components. The term "non-linear" as applied to a characteristic should thus be used in its more general sense when double-input mixers are considered.

The mutual conductance G_1 referred to the first grid may be defined as the partial differential coefficient $\frac{\partial i_a}{\partial v_1}$ (v_3 constant) and partial differentiation of the planar characteristic equation (25) gives

$$G_1 = q = \text{constant} \dots\dots\dots (26).$$

Differentiation of (21) will of course give (20) which expresses the variation of slope G_1 with v_3 . Thus it may be said that for frequency conversion to take place the mutual conductance G_1 must be varied at the

oscillator frequency. That this applies equally well to the single-input mixer can be seen from Fig. 689 where the working point swings at the oscillator frequency between the limits A and B, i.e. between one point where the mutual conductance is low and another where it is high.

From (24) the conversion conductance G_c for the double-input mixer is $\frac{1}{2} a_1 \hat{v}_o$ which again depends on the oscillator amplitude. Under the conditions assumed here the maximum G_c is obtained when the third grid voltage just swings between zero and cut-off. The variation of G_1 with time is then as shown in Fig. 693 and this may be called Class A Operation.

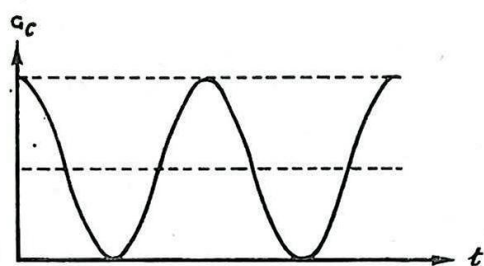


Fig. 693 - Class A operation of double input mixer.

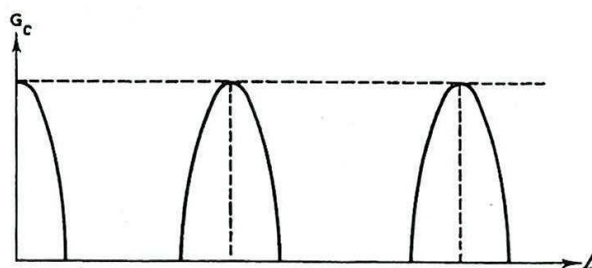


Fig. 694 - Class C operation of double-input mixer.

It is found that a larger conversion conductance can be secured if the valve is driven beyond third grid cut-off during the oscillator cycle. This may be called Class C Operation, and is illustrated in Fig. 694. In this case the variation of G_1 with time is no longer sinusoidal, but since it is periodic at the oscillator frequency it may be expressed as a Fourier series :-

$$G_1 = g_0 + g_1 \cos \omega_o t + g_2 \cos 2\omega_o t + g_3 \cos 3\omega_o t + \dots, \dots (27).$$

With this substitution, (19) becomes

$$i_a = (g_0 + g_1 \cos \omega_o t + g_2 \cos 2\omega_o t + g_3 \cos 3\omega_o t + \dots) (V_1 - V_c + \hat{v}_s \cos \omega_s t) \dots \dots \dots (28),$$

where V_1 is the bias on the first grid. When (28) is expanded it contains product terms of the form :-

$$g_n \hat{v}_s \cos \omega_s t \cdot \cos n\omega_o t; \quad (n = 1, 2, 3, \text{etc.}).$$

$n = 1$ will give the first order mixing components, $n = 2$ the second order and so on. With $n = 1$ the conversion conductance for first order mixing is found to be :-

$$G_c = \frac{1}{2} g_1 \dots \dots \dots (29).$$

The value of g_1 depends on the bias and the amplitude of the oscillator voltage applied to the third grid. If this bias is obtained by clamping at this grid (see Chap. 12), so that the third grid voltage never rises appreciably above zero, the first order conversion conductance varies with the amplitude of the oscillator voltage in the manner shown in Fig. 695.

For increasingly higher orders of mixing the conversion conductance falls off rapidly unless the oscillator voltage is of very large amplitude. Measurements taken under typical operating conditions with a reasonable value of oscillator voltage gave relative values of 12, 6 and 1 for the conversion conductances for first, second and third order mixing respectively.

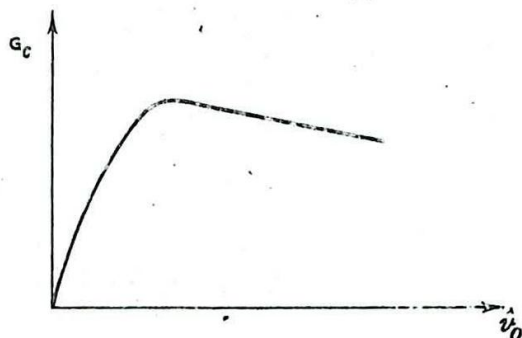


Fig.695_ First order mixing : variation of conversion conductance with amplitude of oscillator voltage.

19. The Two-Pole Converter

Conventional frequency converters are not in general suitable for radar equipments operating at signal frequencies above about 200 Mc/s. For conversion at these frequencies the conversion element is a diode or a crystal; i.e. a two-pole converter.

The problem of analysis of the two-pole converter is considerably complicated by the fact that the signal frequency input, local oscillator input, converter element and IF output load are all in series. There may also be a steady voltage source in the circuit to provide bias; this is shown as a battery in the basic circuit of Fig. 696. The assumption of independence of input and output no longer holds, but in order to reduce the problem to manageable proportions it is necessary to make certain simplifications, namely :-

- (1) the signal is derived from a circuit the impedance of which is zero at all frequencies not close to the signal frequency, where it may be regarded as resistive;
- (2) the impedance of the IF circuit is similarly resistive close to the intermediate frequency and zero for all other frequencies;
- (3) the effective local oscillator voltage is assumed to be injected from a source of negligible impedance at all frequencies;
- (4) the amplitude of the applied local oscillator voltage is very much greater than the amplitude of either the signal or IF voltages;
- (5) the susceptance of the converter element is zero.

The above assumptions naturally limit the general application of the analysis, but to a first approximation they appear to give a satisfactory explanation of the conversion process. The equivalent circuit is shown in Fig. 697; it should be remembered that R_s exists only for the signal frequency and R_i only for the intermediate frequency.

Suppose that the oscillator voltage is applied without any signal; then there will be no voltage drop across R_s , which is effective at the signal frequency only. Since there is no signal input no IF

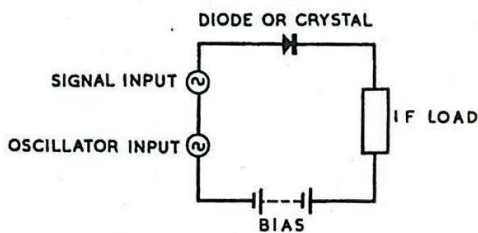


Fig. 696 -
Basic circuit of two-pole
converter.

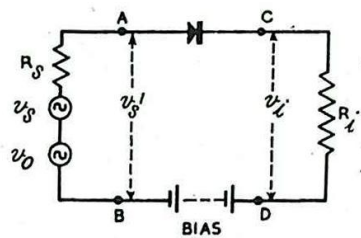


Fig. 697 -
Equivalent circuit of two-
pole converter.

component is present in the circuit so that no voltage is developed across R_i . The waveform of the current through the converter element is then obtained by the usual construction of Fig. 698. One cycle only is shown in this figure; several cycles are drawn in Fig. 699. If the equation of the converter characteristic is :-

$$i_a = f(v) \dots\dots\dots (30),$$

and the voltage v at any instant is

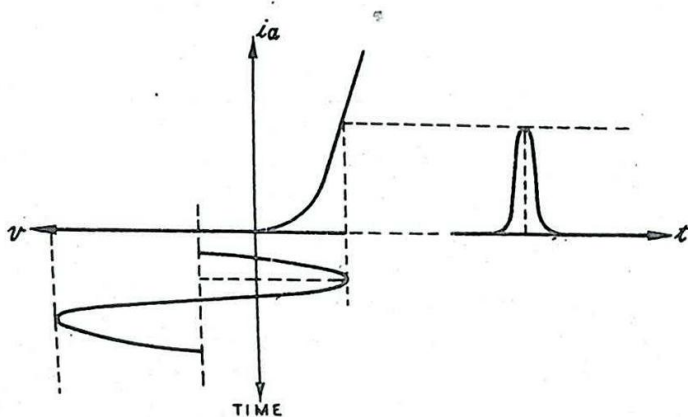


Fig. 698 - Current pulse in converter circuit when oscillator voltage and bias only are applied. (one cycle only shown).

$$v = V_0 + \hat{v}_0 \cos \omega_0 t \dots\dots\dots (31),$$

the instantaneous current is

$$i_a = f(V_0 + \hat{v}_0 \cos \omega_0 t) \dots\dots\dots (32).$$

This is the equation of the current waveform of Fig. 699(a) and can be expressed in the form of a Fourier series :-

$$i_a = I_0 + I_1 \cos \omega_0 t + I_2 \cos 2\omega_0 t + I_3 \cos 3\omega_0 t + \dots\dots (33),$$

where the coefficients $I_0, I_1, I_2 \dots\dots$ etc., can be calculated given the characteristic equation (30) and V_0 and \hat{v}_0 .

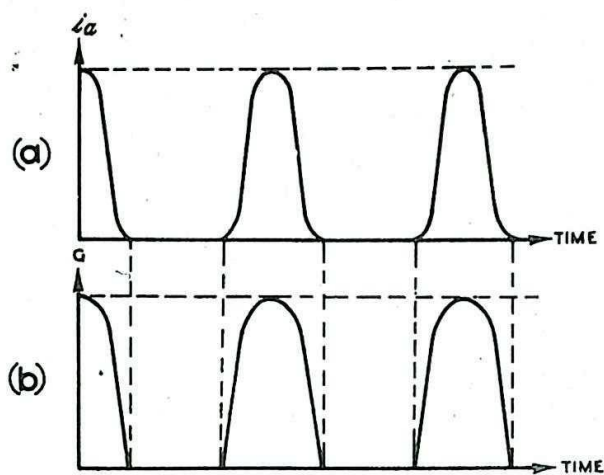


Fig.699 - Variation of current (a) and conductance (b), for two-pole mixer when oscillator voltage and signal only are applied.

The conductance of the converter element is defined as :-

$$G = \frac{di_a}{dv} = f'(v) \dots\dots\dots (34),$$

and therefore the instantaneous conductance is, from (31) and (34)

$$G = f'(V_o + v_o \cos \omega_o t) \dots\dots\dots (35).$$

This also may be expanded as a Fourier series :-

$$G = G_o + G_1 \cos \omega_o t + G_2 \cos 2\omega_o t + G_3 \cos 3\omega_o t + \dots\dots (36).$$

The instantaneous conductance is plotted against time in Fig. 699(b).

It is now necessary to consider what happens when the signal input is applied. Firstly, the signal voltage gives rise to a current in the converter circuit, (Fig. 697), which, since the converter element is a non-linear device, consists of a DC component together with fundamental and harmonic components of the signal frequency. There is a voltage drop across R_s due to the fundamental component only, so that the voltage v_s' actually applied to the converter is of smaller amplitude than the open circuit signal voltage v_s . The applied signal voltage may thus be written :-

$$v_s' = \hat{v}_s' \cos \omega_s t \dots\dots\dots (37).$$

Secondly, the presence of signal and oscillator voltages implies frequency conversion so that there is an IF component in the circuit which produces a voltage across R_i . Let this voltage be v_i so that :-

$$v_i = \hat{v}_i \cos \omega_i t \dots\dots\dots (38).$$

The difference δv between the instantaneous voltage applied to the converter in the signal and no-signal conditions is thus :-

$$\delta v = \hat{v}_s' \cos \omega_s t - \hat{v}_i \cos \omega_i t \dots\dots\dots (39).$$

When the instantaneous voltage is altered by a small amount δv the resultant change in current δi_a is from (34) :-

$$\delta i_a = G \delta v \dots\dots\dots (40),$$

so that the current when the signal, as well as the oscillator voltage, is applied is

$$i_a = f(v) + G \delta v \dots\dots\dots (41).$$

$f(v)$, G and δv are given by (33), (35) and (39) respectively; therefore, on substituting, we obtain :-

$$\begin{aligned} i_a = & I_0 + I_1 \cos \omega_o t + I_2 \cos 2\omega_o t + I_3 \cos 3\omega_o t + \dots\dots\dots \\ & + (G_0 + G_1 \cos \omega_o t + G_2 \cos 2\omega_o t + G_3 \cos 3\omega_o t + \dots\dots\dots) \\ & (\hat{v}'_s \cos \omega_s t - \hat{v}_i \cos \omega_i t) \dots\dots\dots (42). \end{aligned}$$

If this is expanded, the only terms of interest are those giving the signal frequency and intermediate frequency currents, and for first order mixing these are :-

$$(G_0 + G_1 \cos \omega_o t) (\hat{v}'_s \cos \omega_s t - \hat{v}_i \cos \omega_i t) \dots\dots\dots (43).$$

Since $\omega_i = \omega_s - \omega_o$, rearrangement of (43) gives the amplitude i_s of the signal frequency current :-

$$\hat{i}_s = G_0 \hat{v}'_s - \frac{1}{2} G_1 \hat{v}_i \dots\dots\dots (44),$$

and the amplitude \hat{i}_i of the intermediate frequency current :-

$$\hat{i}_i = \frac{1}{2} G_1 \hat{v}'_s - G_0 \hat{v}_i \dots\dots\dots (45).$$

To (44) and (45) must be added a third relation :-

$$\hat{v}_i = R_i \hat{i}_i \dots\dots\dots (46).$$

These three equations enable the performance of the mixer element to be determined completely, but it should be remembered that the simplicity of the relations is apt to be misleading. G_0 and G_1 must be obtained by Fourier analysis, and this cannot be done unless the characteristic equation of the element, the bias and the amplitude of the oscillator input are all known. For n th order mixing it is easy to show that the only difference in (44) and (45) is the substitution of G_n for G_1 :

Despite the mathematical complication of a complete solution, some general results may be obtained. If (44), (45) and (46) are solved for \hat{i}_s and \hat{i}_i the results are :-

$$\hat{i}_s = \left\{ G_0 - \frac{1}{4} \frac{G_1^2 R_i}{1 + G_0 R_i} \right\} \hat{v}'_s \dots\dots\dots (47)$$

$$\hat{i}_i = \frac{1}{2} \left\{ \frac{G_1}{1 + G_0 R_i} \right\} \hat{v}'_s ; \dots\dots\dots (48),$$

but the input impedance of the mixer at the signal frequency, i.e. the impedance seen when "looking in" at AB of Fig. 697 is \hat{v}'_s / \hat{i}_s and this is

given directly by (47). Its value will thus depend on R_i , the load resistance at the intermediate frequency. Similarly it may be shown that the IF output impedance, i.e. the impedance seen "looking back" into CD depends on R_s , the impedance of the signal frequency source. This is implied in (48) where \hat{i}_i is seen to be proportional to v_s' and this in turn depends on the signal frequency voltage drop across R_s .

Although the mixer element is non-linear, equations (44), (45) and (46) are all of the first degree so that it is possible to apply the concept of equivalent circuits to the mixing process. It is important to note that this holds only for small signal inputs (see equation (40)). It is found that the same equivalent circuit holds for both signal and IF, viz, that of Fig. 700; (G_s and G_i are the conductances of the signal and IF circuits respectively). In this circuit the generator v_s must be considered to be at the signal frequency when the input signal current is being calculated and at the intermediate frequency for determining the IF output current. Its magnitude is the same in both cases.

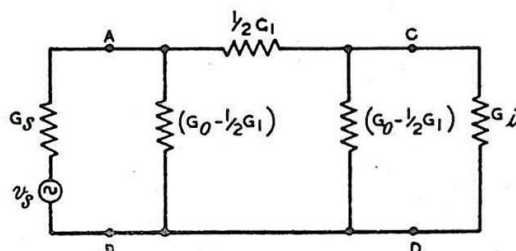


Fig. 700 -
Equivalent circuit of two-pole mixer.

Since the mixer draws current from the signal source it is preferable to define the conversion process in terms of power. For a given value of \hat{v}_s the IF

power delivered to G_i is a maximum when both G_i and G_s

are equal to the characteristic admittance of the π -section attenuator between AB and CD; (See Chap. 3 Sec. 4). The formula given in Chap. 3 Sec. 7 for the characteristic conductance reduces, after the appropriate change of symbols, to

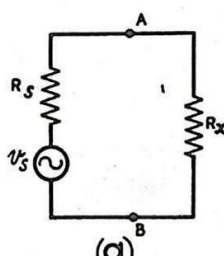
$$G_s = G_i = \sqrt{G_0^2 - \frac{1}{4} G_1^2} \dots\dots\dots (49).$$

This matched condition does not necessarily give the best signal/noise ratio in the receiver, so that a more general approach to the problem is sometimes desirable.

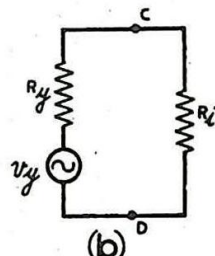
For the signal frequency, the part of the circuit to the right of AB in Fig. 700 may be replaced by a

single conductance (G_x) or resistance (R_x) giving the signal frequency equivalent circuit of

Fig. 701(a). Similarly, by Thevenin's Theorem, when the IF is being considered the part to the left of CD may be replaced by a generator v_y in series with the output resistance R_y giving the IF equivalent circuit of Fig. 701(b). The peak signal and IF currents in



(a) SIGNAL FREQUENCY



(b) INTERMEDIATE FREQUENCY

Fig. 701 -
Equivalent circuits of two-pole mixer.

these two circuits are \hat{i}_s and \hat{i}_i respectively, and the Power Conversion Loss of the mixer may be defined as $10 \log A$ db, where

$$A = \frac{\text{Power in signal frequency equivalent circuit}}{\text{Power in IF equivalent circuit}}$$

$$= \frac{\hat{i}_s^2 (R_s + R_x)}{\hat{i}_i^2 (R_y + R_i)} \dots\dots\dots (50).$$

(Several alternative definitions exist, usually implying a matched condition. The definition given above covers all these and has the advantage of being of general application).

The value of A will always be greater than unity but may approach it in certain cases. Referring to Fig. 700, if the conductance $(G_0 - \frac{1}{2} G_1)$ of the shunt arms could be reduced to zero then the two equivalent circuits of Fig. 701 would become identical and the power in each would be the same. Unfortunately it is not possible to reduce the difference between G_0 and $\frac{1}{2} G_1$ to zero without at the same time making them zero individually so that the series arm of Fig. 700 also becomes zero. There will thus be no IF output unless G_1 is also zero, i.e., the dynamic resistance of the IF load circuit must be infinite. Thus, while 100% power conversion is theoretically possible, it is unattainable in practice, the position being analogous to the case of the Class C amplifier, which can theoretically attain 100% efficiency but only if the load impedance in its anode circuit is infinite.

At this point it is convenient to treat the diode and crystal separately, there being important differences between them which will be emphasised by the examination of each in turn.

20. The Diode Mixer

A possible circuit for a diode mixer is shown in Fig. 702. L_1 and C_1 form the anode load of the RF stage and L_1 is tapped for connection to the diode. There is thus a step-down auto-transformer between the RF anode and the diode input. The local oscillation is injected into the tuned circuit through the small capacitance C_2 , and it is assumed that the IF is low enough to permit an oscillator voltage of adequate amplitude to be applied to the diode by this means. T is the IF output transformer, C_3 being

connected across the primary to bypass the signal and oscillator frequency currents and their harmonics. The battery of Fig. 45 has been replaced by the resistor R_b shunted by C_4 , the bias being produced by the mean diode current flowing through R_b .

If the behaviour of this circuit is analysed

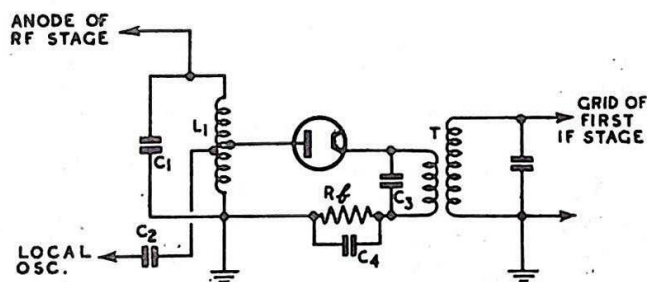


Fig. 702 - Diode mixer circuit.

it will be found that, for a given input to the grid of the RF stage, maximum voltage at the grid of the first IF is secured when both input and output circuits are matched to the characteristic conductance of the π -section of the equivalent circuit of Fig. 700. The position of the tapping point on L_1 and the turns ratio of the IF transformer T will therefore depend on the value of this characteristic conductance.

As a numerical example of the design considerations the simple case of a linear diode will be examined. The diode may be supposed to have a constant conductance G_+ in the forward direction and zero conductance in the reverse direction. The current waveform of Fig. 699 will then become a succession of peaks of a sine wave while the variation of conductance during the cycle will be represented by a square wave. In Fig. 703 anode current and conductance are shown plotted against the phase angle θ of the oscillator voltage. If the conduction angle of the diode is 2ϕ then all the relevant quantities may be expressed in terms of ϕ as a parameter. The results are :-

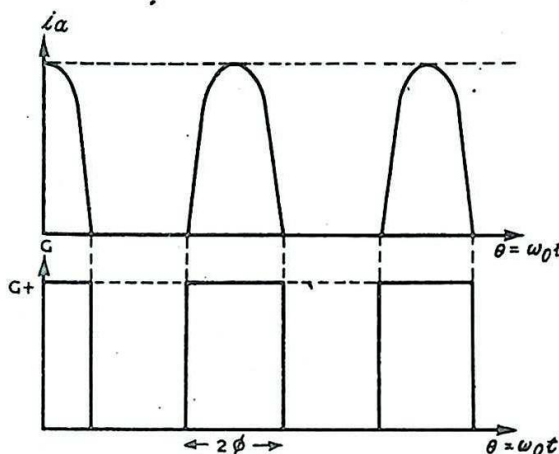


Fig. 703--

Modified form of fig. 699 for the case of a linear diode.

$$\text{Mean diode current } \bar{i}_a = \frac{1}{\pi} \hat{v}_0 (\sin \phi - \phi \cos \phi) G_+ \dots (51).$$

$$\begin{aligned} \text{Bias voltage} = V_b &= R_b \bar{i}_a \\ &= \hat{v}_0 \cos \phi \dots (52) \end{aligned}$$

$$G_0 = \frac{\phi}{\pi} G_+ \dots (53)$$

$$G_1 = \frac{2G_+}{\pi} \sin \phi \dots (54)$$

$$G_0 - \frac{1}{2} G_1 = \frac{G_+}{\pi} (\phi - \sin \phi) \dots (55)$$

$$(G_0^2 - \frac{1}{4} G_1^2)^{\frac{1}{2}} = \frac{G_+}{\pi} (\phi^2 - \sin^2 \phi)^{\frac{1}{2}} \dots (56).$$

These relations apply whether or not the π -section is properly terminated. If in addition this condition is assumed, the four resistances R_s , R_x , R_y

and R_1 of equation (50) are all equal and the conversion loss-factor A is then simply the ratio of the squares of the signal and IF currents. Dividing (47) by (48) and putting R_1 equal to the characteristic resistance, the value of A is found to be :-

$$A = \frac{4 \left\{ G_0 + \sqrt{G_0^2 - \frac{1}{4} G_1^2} \right\}^2}{G_1^2}$$

$$= \frac{\left\{ \phi + \sqrt{\phi^2 - \sin^2 \phi} \right\}^2}{\sin^2 \phi} \dots\dots\dots (57).$$

It follows from (51) and (52) that ϕ is independent of the amplitude \hat{v}_o of the local oscillator voltage and depends only on the bias resistance R_b . This is a natural consequence of the assumption of a linear diode characteristic. The values are tabulated below for two values of ϕ : 30° and 60° , the conductances being given in micromhos.

	$\phi = 30^\circ$	$\phi = 60^\circ$
G_o	333	667
G_1	636	1070
$G_o - \frac{1}{2} G_1$	15	132
$\sqrt{G_o^2 - \frac{1}{4} G_1^2}$	98	345
A	1.84 (2.64 db.)	3.58 (5.54 db.)
R_b	29.2 k Ω	2.3 k Ω

The increase in power conversion efficiency by decreasing the conduction angle 2ϕ from 120° to 60° is well brought out by the above table. At the same time it should be noted that the characteristic conductance has decreased from 345 micromhos to 98 micromhos so that higher impedance circuits are needed to take advantage of the increased efficiency. It may be difficult to design these, particularly the signal frequency circuit, especially if this frequency is high, and this fact, rather than the operating conditions of the diode, may limit the obtainable power conversion efficiency.

When the diode characteristic is non-linear, as it is in practice, the conduction angle is no longer independent of the amplitude of the oscillator voltage and the relations taking the place of equations (51) to (57) are more complicated functions of ϕ . Once the circuit has been designed it is essential to keep the operating conditions constant. If the oscillator input varies, the matching is upset and the efficiency falls off. The easiest quantity to measure is the mean diode current, and a jack is connected in series with the bias resistor R_b so that a meter may be plugged into the circuit. For any given circuit there is a definite mean current, and provision is made for varying the coupling between the oscillator and mixer so that this current can be obtained.

21. The Effect of Diode Capacitance and Lead Inductance

In the foregoing analysis it has been assumed that the diode can be represented by a non-linear conductance, the capacitance between the electrodes being neglected. In a practical diode this capacitance will not be negligible and will become increasingly important as the frequency is raised. There are two effects to be considered. In the circuit of Fig. 702, provided C_3 and C_4 are large compared with the

capacitance of the diode they have very little signal and oscillator voltage developed across them and thus the diode capacitance is of little consequence from the point of view of potential division in the series circuit between the tapping on the RF coil and earth. Hence, this series circuit will, at the signal frequency, be a capacitance practically equal to the capacitance of the diode, and since this is connected across part of the signal circuit it may, if too large, limit the dynamic resistance which may be obtained. This will cause a loss of gain in the RF stage.

A more important effect is that of the inductance of the leads to the diode electrodes, so that, in effect, the valve constitutes a network such as that of Fig. 704. Even if in the circuit of Fig. 702 the leads between the diode and the other circuit components may be made so short that their inductance is negligible, the leads inside the diode bulb are still present. A series resonant circuit may then be set up and the diode will present a low impedance to the tapping on L_1 , making it difficult to develop adequate signal voltage at the diode. Reduction of the capacitance of the diode raises the series resonant frequency and extends the useful range of the diode.

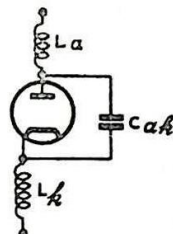


Fig.704-
Effective diode circuit
showing lead inductances.

The static characteristic of a diode may be reasonably well represented by the three-halves-power relation :-

$$i_a = k (v + b)^{3/2} \dots\dots\dots (58),$$

where b is a quantity depending upon contact potentials. For a planar diode with negligible initial electron velocities, the constant k is given by :-

$$k = 2.34 \cdot 10^{-6} \frac{a}{d^2} \dots\dots\dots (59)$$

where a = emitting area of the cathode in cm^2 ,

d = cathode/anode spacing in cm. i_a and v are in amps, and volts respectively.

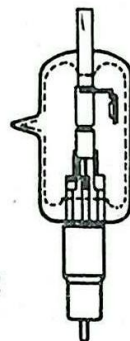
The capacitance (in pF) is given by :-

$$C_{ak} = 8.85 \cdot 10^{-2} \cdot \frac{a}{d} \dots\dots\dots (60).$$

It is important to have as high a value of k as possible since this gives high values of G_0 and G_1 and enables the power conversion to be maintained at high frequencies. At the same time, for the reasons stated above, C_{ak} should be kept as small as possible. Comparison of (59) and (60) shows that for a given value of k , halving the spacing between anode and cathode enables the area of the cathode to be reduced to one quarter, C_{ak} thus being halved. The best diode will then be one with a very small spacing and a correspondingly small cathode area. The effect of lead inductance may be overcome by constructing the diode so that it may be built into, and therefore form an integral part of, a coaxial line circuit. A drawing of the CV58 diode, which has been designed on these principles, is shown in Fig. 705.

22. The Crystal Mixer

The crystal "valve", used primarily as a frequency converter for centimetre wavelengths, is shown in section in Fig. 706. The rectifying contact is made by placing a pointed tungsten "whisker" on a smooth silicon surface, the direction of electron flow being from the tungsten to the silicon. During manufacture the position and pressure of the tungsten whisker are adjusted to give a good "front-to-back ratio" and also a specified forward resistance.



Finally wax is inserted into the crystal capsule through a hole in the ceramic holder. Its functions are to prevent displacement of the tungsten wire from the silicon and to prevent the ingress of moisture.

Fig. 705 - CV 56 diode.

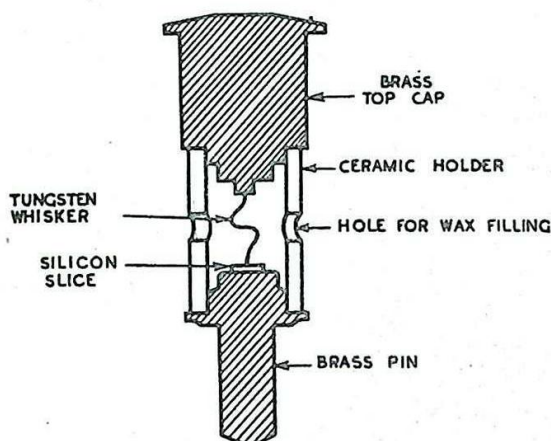


Fig. 706 -
Section of crystal valve.

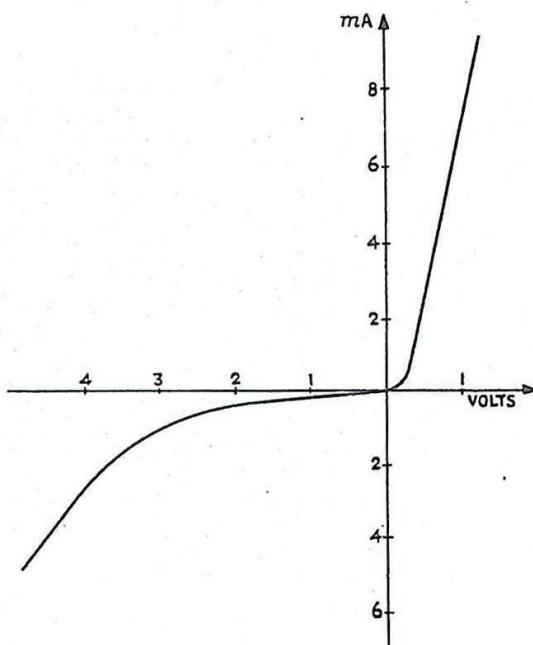


Fig. 707 - Static characteristic of crystal.

The static characteristic of a typical crystal is shown in Fig. 707 and the difference between this and the corresponding curve for a diode is at once apparent. The presence of reverse current modifies the current and conductance waveforms of Fig. 699 to those of Fig. 708, the value of G_0 , the mean conductance over the oscillator cycle, being increased, while G_1 is decreased. The conductance of the shunt arms of the π -network of Fig. 700 is therefore increased while that of the series arm is decreased, both changes leading to a fall in conversion efficiency. This is particularly so when the oscillator voltage at the crystal exceeds one or two volts, causing the reverse current to increase considerably.

It is therefore to be expected that an optimum oscillator input of fairly small amplitude gives maximum conversion efficiency. In fact it is necessary to use an even smaller oscillator input in order to achieve the optimum signal/noise ratio, so that in practice the maximum conversion efficiency is never reached; (see Sec. 27).

Mixer circuits for use with crystals are of two general types, coaxial line and waveguide, and these are shown diagrammatically in Figs. 709 and 710. These designs are made possible by the experimental fact that the resistive component of the RF impedance of the crystal does not vary greatly from one crystal to another. In the coaxial type of mixer circuit the crystal acts as the termination of the line, the stub being connected at a fixed distance from the crystal. The reactive portion can then be tuned out by the stub adjustment so that the input line is correctly terminated. In the waveguide type of mixer

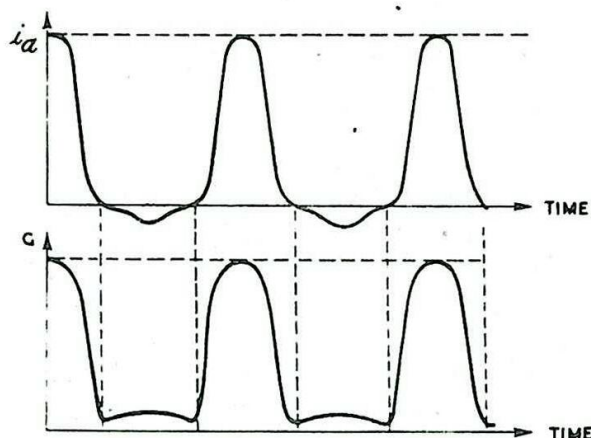


Fig. 708 -

Modified form of fig. 699 showing effect of reverse current.

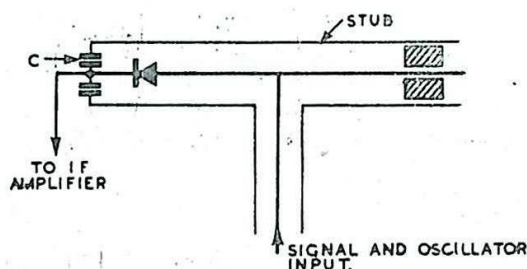


Fig. 709 -

Diagrammatic sketch of coaxial mixer circuit.

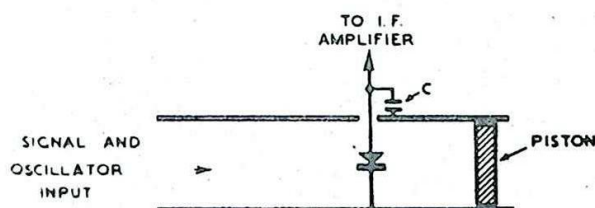


Fig. 710 -

Diagrammatic sketch of waveguide mixer circuit.

circuit the crystal is placed across the guide parallel to the electric field, the guide dimensions being chosen so that the resistive component of the RF impedance is matched to the wave impedance of the guide. The reactive component is tuned out by adjustment of the piston behind the crystal. The RF by-pass capacitance C is in both cases of the "built-in" type, a dielectric washer or metal flange providing all the capacitance that is required.

It is essential that the match between the crystal and the input line or guide should be as good as possible. Any mismatch causes partial reflection at the termination so that part of the signal energy

is reflected back to the aerial and reradiated, producing a falling-off in performance of the equipment.

23. Effect of Crystal Capacitance

Physical investigations of the phenomenon of crystal rectification indicate the existence of a Boundary Layer of the order of 10^{-6} cm. depth just within the silicon block. This layer is responsible for the rectifying action, as it is only therein that the property of non-linearity occurs. It may be shown that there results an effective capacitance associated with this boundary layer which is termed the Contact Capacitance. The presence of this inherent capacitance in direct association with the rectifying layer limits the high frequency performance to be expected from a crystal mixer.

An equivalent circuit for a crystal is shown in Fig. 711.

R_b is the boundary layer resistance which is small in the forward and large in the backward direction, and this is shunted by the boundary layer capacitance C_b . R_s ,

the Spreading Resistance, is the resistance of the silicon block apart from the boundary layer.

R_c includes the resistance of the tungsten whisker and also the contact resistance between the

point of the whisker and the upper surface of the boundary layer. L is the inductance of the whisker. The problem of crystal design is thus to keep the value of C_b small without at the same time increasing the total effective series resistance. Development of crystals for higher frequencies proceeds along these lines.

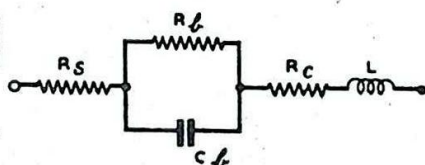


Fig. 711 -

Equivalent circuit of crystal.

Owing to the very small contact area necessary to give a reasonably low value of C_b , crystals are easily damaged by overload, which produces heat at the point of contact and destroys the rectification property. Unlike the diode, which recovers from a momentary overload, the crystal is permanently damaged. For this reason great care should be taken in handling crystals; they should not, in particular, be exposed to stray RF fields, but should be kept in a metal box when not in use.

24. Noise in Frequency Converters

In the frequency converter using conventional types of mixing valves (triode, pentode, hexode) the study of noise follows the same lines as that already given for RF amplifiers in Secs. 7-14. With diode mixers the approach is again along similar lines, but owing to the interdependence of input and output circuits the frequency converter and the subsequent IF stage must be considered together as one composite whole. The same is true of crystal mixers with the added complication that the local oscillator contributes to the noise, this contribution becoming increasingly important as the frequency is raised. The various types of frequency converter are considered in turn.

25. Noise in Triode, Pentode and Hexode Mixers

Since the grid and anode circuits of the valve are tuned to widely different frequencies it is possible to use a triode as a frequency converter without instability occurring. The anode load may be taken to be a pure resistance over its pass-band B so that the ratio of signal to noise power in the output of the converter is equal to the ratio of the

mean square signal and noise currents in the anode circuit. These may be represented by equivalent signal and noise generators in the grid circuit as in the case of the RF amplifier, there being only slight modifications to the argument owing to the frequency conversion process. A simple analysis will be given for the shot noise only. The mean square short-circuit noise current in the anode circuit due to shot noise is given by :-

$$i_n^2 = 2\Gamma^2 \cdot e I_a B \dots\dots\dots (1)$$

where Γ^2 = space charge reduction factor.

e = electronic charge in coulombs.

I_a = anode current in amps.

B = band width in cycles/sec.

In the case of a straight amplifier this noise current may be considered as produced by a fictitious noise generator at the grid of a noiseless valve, the magnitude of the generator voltage being given by

$$\frac{v_n^2}{R} = \frac{2\Gamma^2 e I_a B}{G_m^2} \dots\dots\dots (2).$$

In the case of the frequency converter, I_a and Γ^2 both vary at the oscillator frequency and can no longer be treated as constants. The mean square noise current must therefore be obtained by averaging the expression (1) over the oscillator cycle. Although the effective components of the shot noise are those lying within the IF pass band, the fictitious noise generator at the grid may be considered to be at the signal frequency if the mutual conductance G_m is replaced by the conversion conductance G_c . Equation (2) then becomes :-

$$\frac{v_n^2}{R} = \frac{2e B}{G_c^2} \cdot \overline{(\Gamma^2 I_a)} \dots\dots\dots (3),$$

and the equivalent shot noise resistance R_s is therefore given by :-

$$R_s = \frac{2e B}{K G_c^2} \cdot \overline{(\Gamma^2 I_a)} \dots\dots\dots (4).$$

This can be calculated if the variations of Γ^2 and I_a with grid voltage are known.

The mean square noise current in the anode circuit of a pentode due to both shot and partition noise, is given by :-

$$\overline{i_n^2} = 2e B I_a \left(\frac{I_a + \Gamma^2 I_a}{I_a + I_s} \right) \dots\dots\dots (5),$$

where I_s is the screen current. If it is assumed that the ratio of anode to screen currents is constant during the oscillator cycle and equal to m (a not unreasonable assumption) equation (5) then becomes :-

$$\overline{i_n^2} = 2 e B I_a \left(\frac{1 + m \Gamma^2}{1 + m} \right) \dots\dots\dots (6).$$

This must be averaged over the oscillator cycle and the result divided by $K G_c^2$ to give the equivalent noise resistance R_e .

Equation (6) may be re-written

$$\frac{1}{I_n^2} = 2 e B I \left\{ \frac{m (1 + m \Gamma^2)}{(1 + m)^2} \right\} \dots\dots\dots (7),$$

where I is the total cathode current ($I_a + I_s$).

A hexode operates with a cathode current which is practically constant, the third grid controlling the division of current between second grid (screen) and anode. I and Γ^2 may therefore be taken as constant while m varies at the oscillator frequency. It is then found that the average value of (7) taken over the oscillator cycle is considerably higher than either (4) or (5). That this should be so follows from first principles since one method of reducing partition noise is to increase the ratio of anode to screen currents; (see Chap. 15 Sec. 9.). Anything which reduces this ratio, such as the application of a negative voltage to the third or injector grid, therefore increases partition noise.

From the above arguments it would be expected that a pentode frequency converter would be noisier than a triode and a hexode noisier than a pentode, and this in fact is found to be the case. It is found also that a given valve is noisier as a frequency converter than as a straight amplifier, the mean anode current being the same in both cases; this is mainly due to the lower value of G_c compared with G_m . The table below gives comparative figures for the CV660 valve :-

OPERATION	G_m (MILLIMHO)	G_c (MILLIMHO)	EQUIVALENT NOISE RESISTANCE
Straight Amplifier	9.0	-	720 Ω
Frequency Changer (Triode Connected)	-	4.2	1 k Ω
Frequency Changer (Pentode Connected)	-	3.6	3.5 k Ω

Hexode valves have much higher equivalent noise resistances, the 6SA7, for example, giving a value of 220 k Ω .

The actual noise factors obtainable in frequency changers of the conventional type are not of vital importance as these are always used after a signal frequency amplifier where the signal level, and with it the noise produced in the first signal frequency stage, has been raised to such a value that the noise contribution of the mixer itself is of little account. What is of importance is the number of signal frequency stages needed to do this, which can be decided only if the noise factor of the mixer is known. With careful design a triode frequency changer can be made to have a noise factor of about 11 db at 200 Mc/s; this should be compared with 7.6 db and 5.6 db for the CV66 and CV88 respectively operating as straight amplifiers at the same signal frequency; (see Sec. 14.).

26. Noise in Diode Mixers

When discussing the noise present in a diode mixer circuit (see Fig. 7C2) the mixer and the first IF stage must be considered together. The noise fluctuations in the anode circuit of this stage are then due to the following causes :-

- (1) the noise current in the anode circuit of the last RF stage; this includes the inherent noise of this stage together with that generated in the earlier stages and in the aerial, and subsequently amplified.
- (2) The thermal noise of the tuned circuit feeding the diode.
- (3) The shot noise in the diode itself.
- (4) The thermal noise in the IF circuits coupling the diode to the first IF stage.
- (5) The shot, partition and transit-time noise of the first IF stage.

It is possible to derive equivalent circuits for the combination of diode mixer and IF stage by methods similar to those adopted for the RF amplifier in Secs. 7 - 14, but the process is more complicated and only a general outline is given here.

The factors which may be adjusted to give the best signal/noise ratio in the IF output are :-

- (1) The coupling between the signal frequency circuit and the diode.
- (2) The operating conditions of the diode; i.e. bias resistance and local oscillator input.
- (3) The coupling between the diode and the IF stage.

At the frequencies at which diodes are used as mixers it is always possible to obtain a gain in the signal frequency stages which is sufficient to over-ride the noise contribution of the mixer, so that, having chosen the operating conditions of the diode, factors (1) and (3) are then adjusted to give maximum signal output. This leads to the matched condition which is discussed in Sec. 21. The overall noise factor of the combined diode and IF stages can then be calculated or measured. It has been found that for any given bias resistor (R_b of Fig. 702) there is an optimum amplitude of the

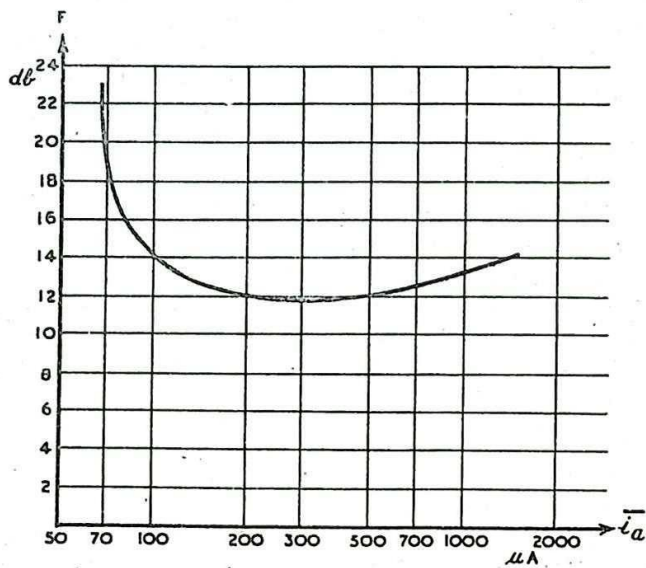


Fig. 712 -
Variation of noise factor with mean diode current.

oscillator input. This is shown in Fig. 712, in which the noise factor for a bias resistor of $5\text{ k}\Omega$ is plotted against mean diode current for a CV58 diode followed by a CV1091 IF stage, the signal frequency being 600 Mc/s and the IF 45 Mc/s . The high value of noise factor for low diode currents is due to the high conversion loss for low oscillator inputs. As the oscillator voltage is increased the conversion loss falls but the shot noise of the diode becomes increasingly important as the mean current rises. For high mean currents the increase in shot noise more than counterbalances the decrease in conversion loss.

There is also a best value of bias resistance, which becomes smaller as the frequency increases; (Fig. 713). This reduction in R_b is necessary to offset the increase in the susceptance due to the diode capacitance. When both oscillator input and bias resistance have been adjusted to their optimum values the resultant minimum noise factor varies with frequency in the manner shown in Fig. 714. The CV58 diode has been used at frequencies up to $3,000\text{ Mc/s}$ in applications where its greater freedom from overload damage is more important than the lower noise level of the crystal. For this diode the average noise factor is about 4 db higher than for a crystal, but manufacturing tolerances for the diodes are less rigid, variations between different diodes being greater than in the case of crystals.

27. Noise in Crystal Mixers

In radar receivers designed for centimetre wavelengths it is not practicable to amplify at the signal frequency so that the mixer must be the first stage in the receiver. Crystals are almost invariably used, and the signal-to-noise ratio of the receiver, and therefore the overall performance of the equipment, depend to a very great extent on

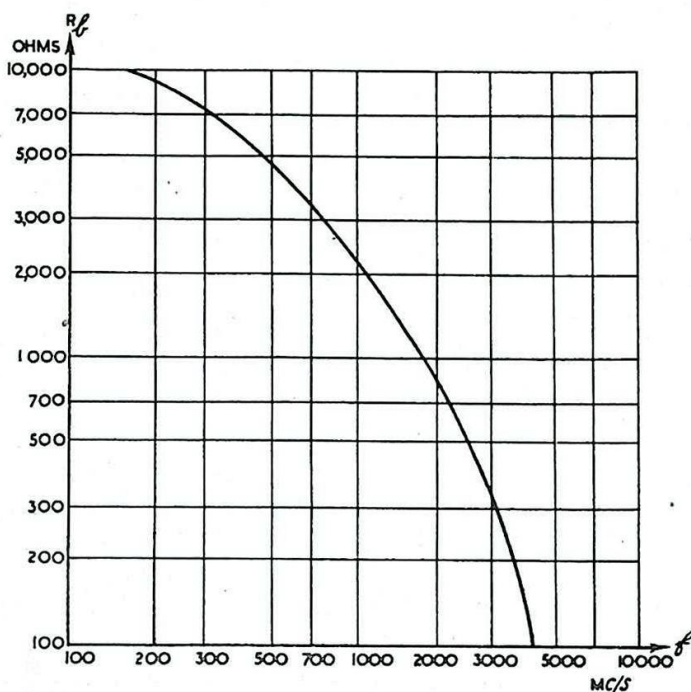


Fig. 713—
Diode mixer: variation of optimum bias resistor with signal frequency.

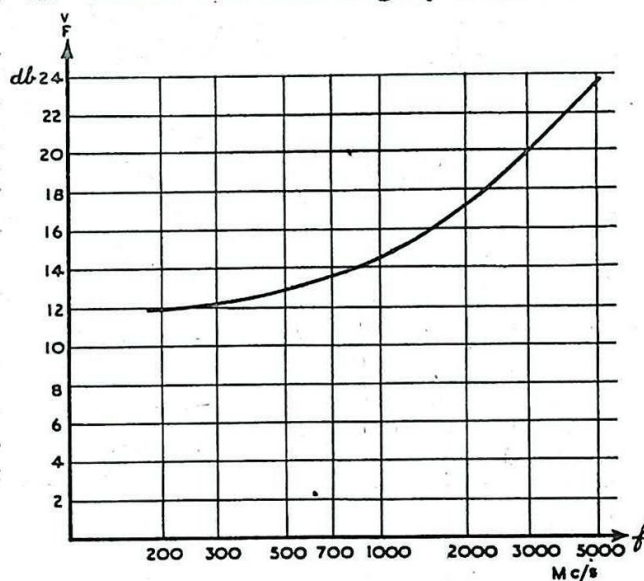


Fig. 714—
Diode mixer: variation of minimum noise factor with signal frequency.

careful design of the crystal and associated circuits for minimum noise.

With the diode, two factors have to be adjusted to give the best performance; the bias resistor and the amplitude of the local oscillator input. With the crystal it is found that no bias resistor is needed, the only variable therefore being the local oscillator input which, as in the case of the diode, is most conveniently represented by the mean rectified current

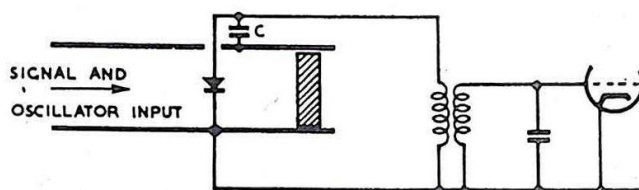


Fig. 715 -
Basic circuit of crystal mixer and IF stage.

which it produces. The crystal is assumed to be matched to the coaxial line or waveguide as described in Sec. 22, so that all the signal power received by the aerial is fed into it. The crystal then acts as a generator at the intermediate frequency, feeding the grid of the first IF stage through a step-up transformer (Fig. 715). It can, in other words, be regarded as replacing the aerial of Sec. 9. The aerial resistance (Fig. 669) must then be replaced by the IF output impedance of the crystal, i.e. by the resistance R_y of the equivalent IF circuit of Fig. 701(b). There will of course be some susceptance in shunt with this due mainly to the RF bypass capacitance C of Figs. 709 and 710 but this will be accommodated by the tuning of the grid circuit of the IF stage.

Owing to conversion loss in the crystal the power developed in the equivalent IF circuit is less than the signal power P_s fed into the crystal. In fact

$$P_s = A \frac{\overline{v_y^2}}{(R_y + R_i)} \dots\dots\dots (8)$$

where v_y , R_y and R_i refer to Fig. 701(b) and A is the conversion loss-factor as defined by equation (50) of Sec. 19. R_i is the load on the crystal imposed by the IF input circuit, and includes the dynamic conductance of the tuned circuit, the transit time conductance and the input conductance of the IF valve due to inductance in the cathode lead. The equivalent input circuit for the IF stage is therefore obtained from Fig. 679 by the simple replacement of v_s by v_y , and t_r^G by t_y^G , it being remembered that t_y^G now refers to the transferred IF output conductance of the crystal; (Fig. 716).

When the noise circuit is considered a complication arises owing to the fact that the noise output of the crystal is higher than that corresponding to the thermal noise in a resistor R_y , i.e.

$$\overline{v_n^2} > 4 k T B. R_y$$

We may therefore write

$$\begin{aligned} \overline{V_n^2} &= x \cdot 4 k T B \cdot R_y \\ &= 4 k (xT) B \cdot R_y \\ &\dots\dots\dots (9), \end{aligned}$$

where x is called the Noise Temperature Ratio of the crystal. Equation (9) shows that an ohmic resistor of value R_y would have to

be raised from room temperature T to a higher temperature xT in order to generate the same noise as the crystal. The equivalent noise circuit of Fig. 680 must therefore be modified by increasing the noise current generator representing the aerial noise by the factor x, giving the circuit of Fig. 717. These equivalent circuits have been drawn assuming that the first IF stage is a pentode, but the case of a grounded grid triode in this position may be dealt with in a precisely similar manner.

Before proceeding to a discussion on the overall noise factor of the crystal and first IF stage it is necessary to examine the factors controlling the noise temperature ratio. The sources of noise are :-

- (1) noise generated in the aerial system and fed into the crystal with the signal;
- (2) noise generated in the crystal itself;
- (3) noise produced by the local oscillator and fed into the crystal with the local oscillation.

For the present (3) is neglected. Aerial noise also is unimportant, as, owing to the high conversion loss of crystals, only a fraction of this noise power is developed in the equivalent IF circuit, and measurements have not shown any detectable contribution of aerial noise to the noise temperature ratio.

The problem of the noise generated in the crystal itself is not fully understood. The mechanism of thermal agitation in the barrier layer is much different from that in a homogeneous metal, and further complication arises because the conditions in the crystal are varying rapidly as the working point is moved up and down the non-linear characteristic by the local oscillator input. There appears to be no correlation between the static characteristic of the crystal and the thermal noise generated in it, and when the question is approached practically by taking measurements of conversion loss and noise temperature ratio there is again no correlation, since a low conversion loss

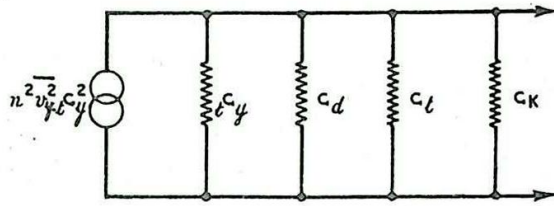


Fig. 716 -
Equivalent IF input circuit of fig. 715.

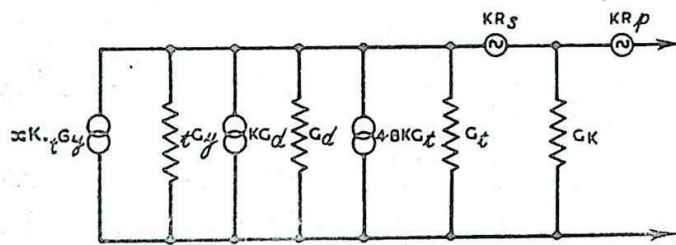


Fig. 717 -
Equivalent noise input circuit of fig. 715.

does not imply a low noise temperature ratio.

The IF output conductance, conversion loss and noise temperature ratio all depend on the mean crystal current \bar{I}_0 . Experimental results for an average crystal operating on about 10,000 Mc/s are shown in Figs. 718 and 719.

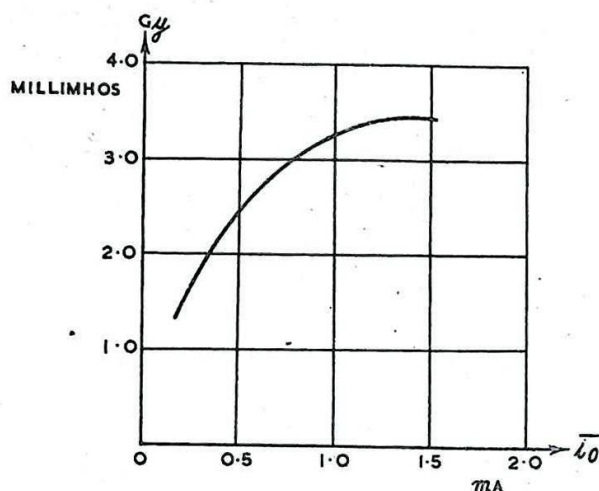


Fig. 718 -
Variation of IF output conductance of crystal with mean crystal current.

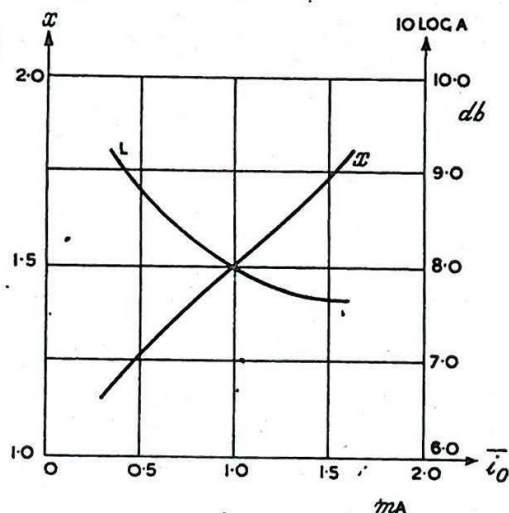


Fig. 719 -
Variation of noise temperature ratio and conversion loss with mean crystal current.

28. Local Oscillator Noise

The local oscillator used in a radar receiver operating on centimetre wavelengths is usually a klystron, and this type of valve produces, in addition to the desired oscillation of frequency f_0 , random noise which is most intense at frequencies close to f_0 . The amplitude spectrum of the output of the klystron is indicated in Fig. 720. The amplitude of the noise relative to the desired oscillation is, of course, very much smaller than is shown in the figure. To understand how this noise is produced, consider the double rhumbatron klystron shown diagrammatically in Fig. 721. The local oscillator (Sutton Tube) is of the reflector, or single-rhumbatron type, but the mechanism of noise production is similar and the explanation is more easily followed with the double-rhumbatron type. The operation of both double-rhumbatron and reflector type klystrons is explained in Chap. 8 Secs. 21-23.

The mechanism of noise production is as follows. The electron beam entering the buncher at A is not uniform but is intensity modulated in a random manner due to the shot noise fluctuations of the emission from the cathode. The buncher then

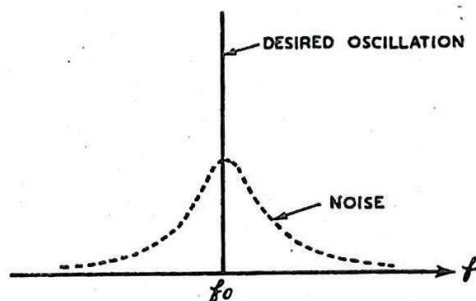


Fig. 720 -
Amplitude spectrum of output of klystron local oscillator.

acts as a catcher for the shot noise fluctuations and noise current will thus be induced in it in a manner similar to that in which induced grid noise currents are set up in a negative-grid triode. These noise currents are relatively large since the electrons are travelling at high speed, and it is due to them that klystrons are too noisy to be used with advantage as amplifiers for small signal voltages.

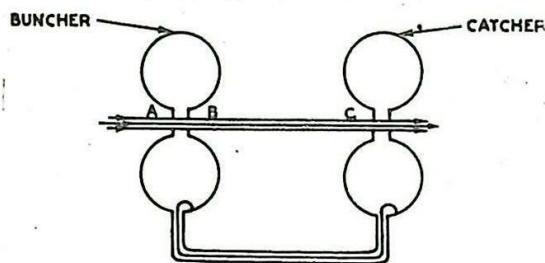


Fig. 721 -
Double-rhumbatron type of klystron.

The noise voltages generated in the buncher modulate the velocity of the main electron stream and thus appear in an amplified form in the catcher. Further amplification is provided by the regenerative coupling between the rhumbatrons. The shot noise fluctuations have an amplitude spectrum which is uniform over all frequencies but the rhumbatrons are frequency-selective and respond only to frequencies near resonance, i.e. close to f_0 . The resultant noise voltages are therefore largest for frequencies close to f_0 , the amplitude falling off rapidly as the frequency separation from f_0 increases. The general nature of the amplitude spectrum is thus as shown in Fig. 720. As would be expected, the frequency spread of the noise voltages depends on the selectivity of the rhumbatrons, high-Q rhumbatrons giving a small spread and vice versa.

The output circuit of the mixer is tuned to the intermediate frequency and therefore it is necessary to consider only those local oscillator noise components, which, after passing through the crystal, produce an output lying within the pass-band of the IF amplifier. If f_0 is the frequency of the local oscillator, f_i the intermediate frequency and B the band width of the IF amplifier, then the effective local oscillator noise components are those lying in the two frequency ranges.

$$(1) \quad f_0 - f_i \pm \frac{1}{2} B$$

and

$$(2) \quad f_0 + f_i \pm \frac{1}{2} B.$$

It is these which give IF outputs by first order mixing with the local oscillation. The noise produced by mixing between the individual noise components themselves and between noise components and signal is negligible. The relevant frequency bands are shown in Fig. 722.

From what has been said of the fundamental cause of klystron noise it follows that it is essentially a selectivity problem and any factors leading to a low degree of selectivity increase the contribution of local oscillator noise to the noise temperature ratio. The two most important are :-

- (1) a low-Q rhumbatron
- (2) a low value of f_i/f_0 .

It is therefore not surprising that local oscillator noise is very small at 3,000 Mc/s and becomes increasingly important as the operating frequency is raised. At 10,000 Mc/s, with an IF of 45 Mc/s, the CV87

and CV129 klystrons produce a very small amount of oscillator noise, but with the CV720 about half the noise temperature ratio of the crystal is due to the local oscillator, average values of x being 1.5 for the CV129 and 3.0 for the CV720. The higher value for the CV720 is due to the smaller dimensions of its cavity, which thus has a lower Q .

Another point of interest with the CV720 arises from its use in AFC circuits (Chap. 8 Secs. 49-53) where a certain amount of control over the oscillation frequency is obtained by variation of the reflector voltage. Measurements show that if too great a frequency variation is attempted by this means the noise temperature ratio rises sharply, causing a serious deterioration in the signal/noise ratio of the equipment. Typical results are shown in Fig. 723.

At a frequency of 24,000 Mc/s most of the noise output of the crystal is due to the local oscillator, noise temperature ratios up to 10 being common. The comparison between results at this frequency and at 10,000 Mc/s is shown in Fig. 724, which also shows the improvement in noise temperature ratio consequent on the use of a high IF with a corresponding increase of f_i/f_o . The 24,000 Mc/s curves of Fig. 724 are drawn for two different values of crystal current, the higher value of crystal current giving the higher noise temperature ratio. This is to be expected, since increasing the local oscillator input to the crystal also increases the noise input.

With a general tendency towards increasing operating frequencies attention has been focused on methods of reducing the local oscillator noise. Amongst the methods suggested are :-

- (1) the use of a high- Q rhumbatron cavity in the local oscillator.
- (2) the use of a high value of IF.

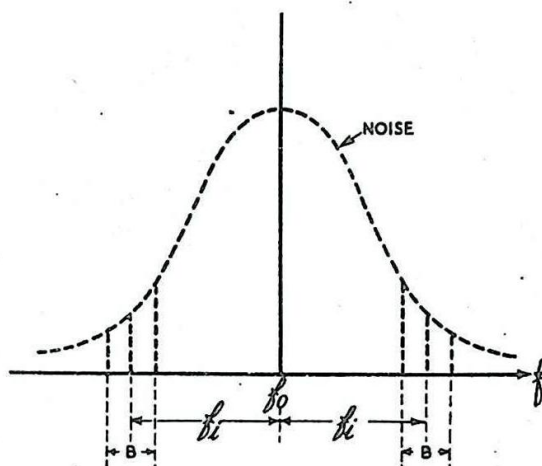


Fig.722-

Amplitude spectrum of output of local oscillator showing noise components producing IF noise in mixer output.

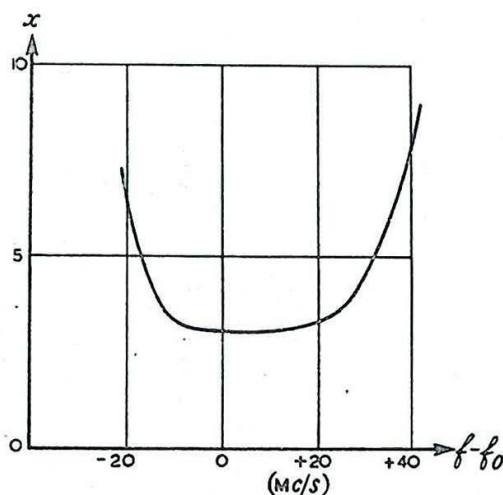


Fig.723-

Dependence of noise temperature ratio on change of frequency produced by variation of reflector volts (90-volt mode of CV 720).

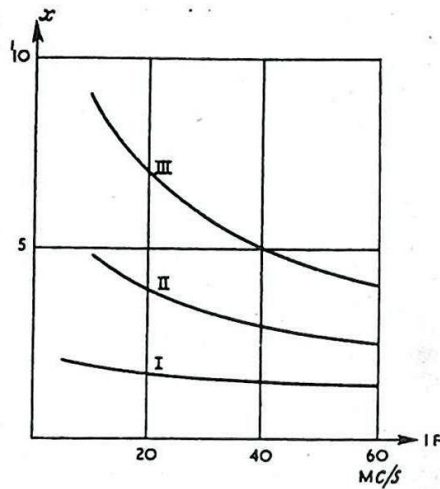


Fig.724 - Dependence of contribution of local oscillator noise on intermediate frequency,

Curve I: 10,000 Mc/s. Mean crystal current = 0.5 mA
 Curve II: 24,000 Mc/s. Mean crystal current = 0.25mA
 Curve III: 24,000 Mc/s. Mean crystal current = 0.5 mA

- (3) filtering the local oscillator output by means of a high-Q cavity between oscillator and mixer.
- (4) eliminating one of the noise-frequency bands of Fig.722. One of these contains the signal and cannot be removed, but the "second channel" can be cut out by suitably positioning the TR cell; (see Sec. 34).
- (5) the use of two mixers in a balanced circuit. The signal outputs are made to add while the noise outputs due to the local oscillator are in antiphase and are therefore cancelled.

29. The Overall Noise Factor of Crystal Mixer and IF Amplifier

As described in Sec. 27, in centimetre receivers it is assumed that the feeder is matched to the crystal so that the aerial resistance R_r of Sec. 8 is replaced by the mixer output resistance R_y of Sec. 19; (see equation (50) of that section and Fig. 701(b)). For an ideal centimetric receiver we assume that all the signal power is supplied to the IF circuit, i.e. that there is no conversion loss, and that the only noise present is that due to thermal noise in R_y .

The equivalent signal and noise circuits for the actual receiver are given by Figs. 716 and 717.

From Fig. 716 the mean square IF voltage at the first grid is given as :-

$$\frac{s_v^2}{G^2} = \frac{n^2 \overline{v_y^2} t_{G_y}^2}{G^2} \quad \text{where } G = t_{G_y} + G_d + G_t + G_K \dots \dots (10)$$

From Fig. 717 the mean square noise voltage at the first grid is given as :-

$$\overline{n^2} = \frac{K(x \cdot t_{Gy} + G_d + 4.8 G_t)}{G^2} + K R_s \frac{(t_{Gy} + G_d + G_t)^2}{G^2} + K R_p \dots (11).$$

Hence, for the actual receiver,

$$\frac{n^P}{s^P} = \frac{\overline{n^2}}{\overline{s^2}} = \frac{K(x \cdot t_{Gy} + G_d + 4.8 G_t) + K R_s (t_{Gy} + G_d + G_t)^2 + K R_p G^2}{n^2 \overline{v_y^2} t_{Gy}^2} \dots (12).$$

For the ideal receiver described above, two modifications to the equivalent circuits are necessary.

(i) Fig. 716. Since there is no conversion loss the mean square signal voltage must be increased by a factor A.

(ii) Fig. 717. Since the only noise present is thermal noise in R_y (Fig. 701(b)) the noise generator $x \cdot t_{Gy}$ must be replaced by t_{Gy} , and the other noise generators must be suppressed.

These modifications are illustrated in the equivalent circuit of Fig. 725. Hence, for the ideal receiver, after making these alterations to equation (12) we obtain :-

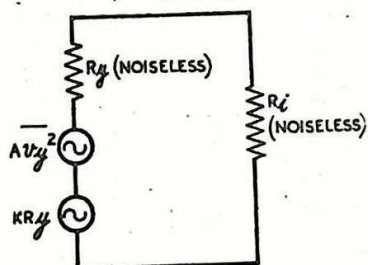


Fig.725 - Equivalent IF input circuit of ideal receiver with crystal mixer.

$$\frac{n^P}{s^P} = \frac{\overline{n^2}}{\overline{s^2}} = \frac{K t_{Gy}}{A n^2 \overline{v_y^2} \cdot t_{Gy}^2} \dots (13).$$

By definition, the noise factor F is the quotient of (12) and (13), viz:-

$$F = A \left\{ x + \frac{G_d}{t_{Gy}} + \frac{4.8 G_t}{t_{Gy}} + \frac{(t_{Gy} + G_d + G_t)^2 \cdot R_s}{t_{Gy}} + \frac{G^2}{t_{Gy}} \cdot R_p \right\} \dots (14).$$

If the noise factor F_i due to the IF amplifier alone is required, it can be obtained directly from equation (24) of Sec. 12 on making the appropriate substitutions. The result is :-

$$F_i = 1 + \frac{G_d}{t_{Gy}} + \frac{4.8 G_t}{t_{Gy}} + \frac{(t_{Gy} + G_d + G_t)^2 \cdot R_s}{t_{Gy}} + \frac{G^2}{t_{Gy}} \cdot R_p$$

..... (15).

A comparison of equations (14) and (15) shows that

$$F = A \left\{ F_i + x - 1 \right\} \quad \text{..... (16).}$$

This result is a particular case of a general theorem for the overall noise factor of several amplifiers or other networks in cascade. If F_1, F_2, \dots, F_n are the noise factors of the individual networks; and M_1, M_2, \dots, M_n the power amplification ratios, the overall noise factor is given by

$$F = F_1 + \frac{F_2 - 1}{M_1} + \frac{F_3 - 1}{M_1 M_2} + \dots + \frac{F_n - 1}{M_1 M_2 \dots M_{n-1}} \quad \text{..... (17).}$$

In the particular case of equation (16), F_1 , the noise factor for the crystal itself, is Ax , and $M_1 = \frac{1}{A}$, while $F_2 = F_i$. With these substitutions (16) is the same as (17) for $n = 2$.

When (15) is differentiated to find the optimum transferred crystal IF output conductance t_{Gy} , the value obtained depends on neither A nor x , for A is merely a multiplying constant and x disappears on differentiation. The minimum overall noise factor F is thus obtained by designing the coupling transformer between the mixer and the first IF stage for a minimum value of F_i .

Since the IF output conductance of the crystal is a function of the mean crystal current (see Fig. 718) it follows that the optimum turns ratio will also depend on the mean crystal current. Once the best operating conditions have been settled it is therefore essential to maintain the correct value of mean crystal current when the receiver is in operation. Failure to do so may cause deterioration in performance.

Values of A and x for an average 10,000 Mc/s crystal are given in Fig. 719. If these values are substituted in equation (16) the results of Fig. 726 are obtained. Two alternative values of F_i , 3 db and 5 db, have been used in calculating these curves. The minimum is due to exactly the same cause as the minimum in the corresponding curve

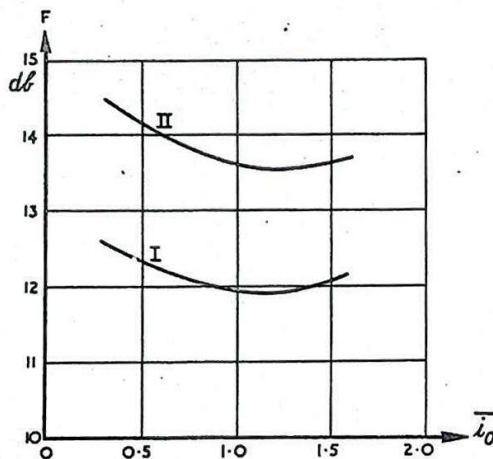


Fig. 726 -

Overall minimum noise factor of IF amplifier and crystal of fig. 720.

$$F = A(F_i + x - 1)$$

Curve I: $F_i = 2(3\text{db})$

Curve II: $F_i = 3.16(5\text{db})$

for the diode (Fig. 712): for small mean crystal currents the conversion loss is high while for large mean currents the crystal noise increases in greater proportion than the decrease in conversion loss.

RADAR RECEIVERS

30. General

In a superheterodyne receiver amplification is obtained in three ways :-

- (1) in the signal frequency stages;
- (2) in the IF stages after frequency conversion;
- (3) in the video stages after detection.

The amplitude of the output from the video amplifier depends on the type of display used. This amplifier must have a band-width sufficient for reasonable pulse reproduction and a gain high enough to give the required output with the given input from the detector.

The detector output, although it is not critical, must lie within well defined limits. A diode is normally the most usual form of detector and this must have a small resistance load if the pulse-shape is to be preserved. Under these conditions the detection efficiency is lower than that common in communication receivers, and the non-linearity of the rectification characteristic at low input levels is more pronounced. With a typical diode (e.g. CV1092) the peak output should not fall much below 1 volt if sensitivity is not to suffer. On the other hand it is not possible to obtain large outputs from the diode without producing overloading in the IF stage preceding it. With the bandwidths commonly used the anode loads of the IF stages must be low (see Chap. 7) and unless the complication of a larger last IF valve is indulged in the detector output begins to be non-linear beyond about 10 volts (peak). The useful output from the detector is thus about 5 volts (peak) and the video amplifier is designed to give the requisite output to the display for this value of input.

It is then necessary to decide how much of the pre-detection gain should be provided by the signal frequency amplifier and how much by the IF amplifier. As a general rule the lower the frequency the easier it is to obtain amplification and maintain stability; moreover, the tuning of the IF amplifier is preset, whereas the signal frequency stages must have provision for varying the tuning to follow alterations in the operating frequency. It is thus desirable to introduce the mixer at the earliest possible place in the amplifier chain. The only reason for using signal frequency amplification at all in radar design is the fact that the noise factor of mixers is in general higher than that of signal amplifiers. If it were possible to produce a mixer with a noise factor as low as that of the signal amplifier the latter could be dispensed with, producing a considerable simplification of the receiver.

31. The Overall Noise Factor of a Receiver

The overall noise factor of a crystal mixer followed by an IF amplifier is dealt with in Sec. 29. The diode and IF amplifier combination also is studied in Sec. 26 and the general problem of RF amplifier noise in Secs. 7-14. In these sections it is assumed that the amplifier stages are linear and the problem is approached by a comparison of the signal and noise powers present at different points. Since signal

and noise are statistically independent sources their mean square values are directly additive, the total power then being the sum of the respective signal and noise powers. It is thus possible to define the overall noise factor of the receiver by comparing the signal and noise powers at the output of the IF amplifier with their values at the aerial or crystal; for although a fundamentally non-linear response is an essential feature of a mixer, nevertheless the mixer equivalent circuits have been shown to be representable to a first approximation by linear circuit elements; (Secs. 15 - 29). Such a simplification is not possible in the case of the video detector, where we are concerned primarily with the envelopes of the signal and noise voltages, and a power concept is not directly applicable. It is therefore common practice to quote the overall noise figure of the receiver as referring to the output of the IF amplifier. From this the minimum received signal power for a "just detectable" signal can be estimated, it being assumed that "just detectable" means equality of signal and noise power at the IF output. A further complication is introduced by the display since it requires a much larger signal to noise ratio at the detector input to produce a clearly defined echo on a PPI than on an A-display. The standard of equal signal and noise powers at the IF output will tend to underestimate the receiver capabilities on a type A-display and to overestimate them if a PPI is used.

32. The Number of Signal Frequency Stages

When amplifier stages are operated in cascade it is possible to deduce the overall noise factor if the noise factors of the individual stages are known. Fig. 76 shows a block diagram of a receiver with the two signal frequency stages preceding the mixer, the latter and the IF amplifier being regarded as one unit in accordance with Secs. 20-29. The overall noise factor up to the IF output is then :-

$$F = F_1 + \frac{F_2 - 1}{M_1}$$

$$+ \frac{F_3 - 1}{M_1 M_2} \dots\dots(1)$$

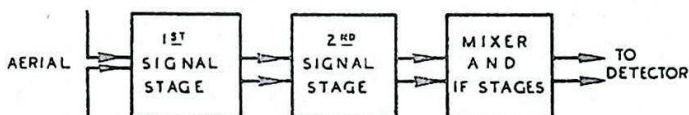


Fig. 727-
Block diagram of receiver with two signal frequency stages.

(see Sec. 29).

where F_1 and M_1 are the noise factor and power amplification respectively of the first signal frequency stage, F_2 and M_2 refer similarly to the second signal frequency stage and F_3 is the overall noise factor of the mixer and IF stages.

As the first numerical example, consider a receiver operating at 30 Mc/s (about the lowest radar frequency). Let F_1 and F_2 be 2 (i.e. 3db) while F_3 is 16, these being reasonable values for CV1136 valves followed by a diode mixer (see Figs. 687 and 714). The power amplification of each signal frequency stage is taken to be 10 so that the overall noise figure from equation (1) is

$$F = 2 + \frac{1}{10} + \frac{15}{100} = 2.25 \dots\dots\dots(2).$$

The number of signal frequency stages has been chosen arbitrarily and it is instructive to examine how the noise factor changes as the number of these stages is altered. With no signal stage the noise factor would, of course, be that of the mixer and IF stages, i.e., 16. With only one signal stage the mixer is now the second stage so that the last term in (1) is omitted and F_2 replaced by F_3 giving

$$F = 2 + \frac{15}{10} = 3.5 \dots\dots\dots (3).$$

With three signal stages,

$$F = 2 + \frac{1}{10} + \frac{1}{100} + \frac{15}{1000} = 2.125 \dots\dots\dots (4).$$

These figures are tabulated below : (Table I)

Number of signal frequency stages	Noise factor F	
	Ratio	db.
0	16	12.0
1	3.5	5.4
2	2.255	3.5
3	2.125	3.3

Two points should be noted: (i) the large improvement (6.6 db) produced by the use of only one signal frequency stage (ii) the very small improvement (0.2 db) when the number of such stages is increased from two to three.

As the second numerical example a signal frequency of 200 Mc/s is chosen, again with CV1136 signal amplifiers preceding the diode frequency converter. From Fig. 687 the noise factor of each signal frequency stage is 11.2 (10.5 db) whereas the noise factor of the diode has not increased appreciably. Assuming that by suitable circuit construction the power amplification can be maintained at its original value of 10, the variation of overall noise factor with the number of signal frequency stages is given in table II.

Table II

Number of signal frequency stages	Noise factor F	
	Ratio	db
0	16	12.0
1	12.7	11.0
2	12.4	10.9
3	12.3	10.9

The improvement in noise factor resulting from the use of one signal frequency stage is now only 1db and the further improvement when more stages are added is negligible. It is clear that the use of CV1136 valves at this frequency is of doubtful benefit, the additional complications of such stages not being balanced by any substantial increase in signal/noise ratio. This is, of course, due to the high noise factor of the CV1136 itself and it must be replaced by a valve of lower noise factor if a substantial improvement is sought. The CV66 has a noise factor of 7.6 db at 200 Mc/s (Fig. 687) which gives overall noise factors of 8.6 db and 8.0 db for one and two stages respectively.

At higher frequencies, say 600 Mc/s, valves such as the CV88 must be used (See Sec. 13). The noise factor of this valve is 10db at 600 Mc/s while that of the diode has risen to 13.4 db (Fig. 714). One signal frequency stage will then give an overall noise factor of 10.8 db

which is reduced by only 0.3 db if a second stage is added. It is therefore sufficient to use one stage preceding the mixer.

33. Receiver for Metre and Decimetre Wavelengths.

The technique employed in the design of radar receivers causes them to fall into two general classes according to the operating frequency; (i) up to 600 Mc/s (ii) above 3000 Mc/s. The upper limit of range (i) approaches the limit of usefulness of valve amplifiers and conventional type oscillators, while the use of waveguides and cavity resonators is not practicable much below the lower limit of range (ii) as they would have to be large and cumbersome. There is a further natural subdivision at about 200 Mc/s, for this represents not only the upper limit of the useful range of the pentode but is also the frequency at which conventional lumped circuits must give way to coaxial lines in which inductance and capacitance are distributed. While the use of coaxial line circuits below this frequency is theoretically advantageous, their use is again precluded on the grounds of excessive bulkiness. Some details of receivers for frequencies up to 600 Mc/s are tabulated below (Table III). These illustrate the points which have been studied in the preceding paragraphs.

TABLE III

Receiver	Operating frequency Mc/s	Number of signal freq. stages	Valve type used in signal amp.	Type of tuned circuit	Valve type used in mixer
A	20-60	3	pentode	lumped	hexode
B	55-84	5	pentode	lumped	diode
C	176	2	pentode	lumped	pentode
D	212	2	triode	lumped	pentode
E	212	2	pentode	coaxial	triode
F	600	1	triode	coaxial	diode

34. Receivers for Centimetre Wavelengths

As signal frequency amplification is not practicable at frequencies of 3000 Mc/s and upwards, the first stage in a receiver operating at such frequencies is of necessity the mixer. A possible arrangement of the signal frequency unit of a receiver operating at about 10,000 Mc/s is shown in Fig. 728. The mixer must be located in the same box as the magnetron if a long and cumbersome waveguide connection between the TR circuits and the mixer is to be avoided. The IF output impedance of the mixer is of the order of 300 ohms (see Sec. 22) and this should be the impedance presented at the connection to the tuned circuit at the grid of the first IF stage.

It is therefore not possible to use a long length of cable to connect the mixer to this stage, for the characteristic impedance of such a cable is never far different from 80 ohms and with such a mismatch at the mixer end a considerable susceptance would be present at the connection to

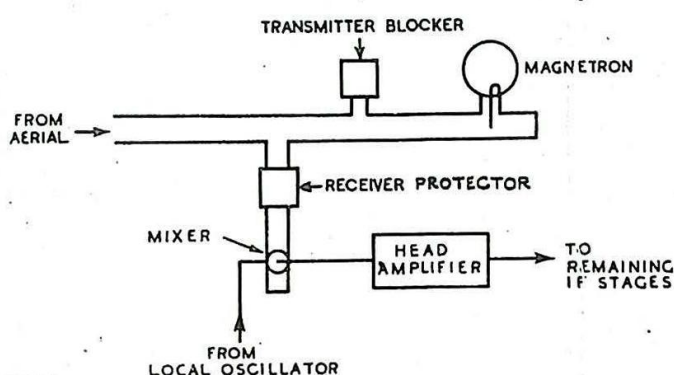


Fig. 728—
Block layout of signal frequency unit of 10,000 Mc/s receiver.

the IF tuned circuit. It is also impracticable to use a step-down transformer at the mixer end of the cable since the extra losses present in such a transformer would be an additional source of noise. It is therefore standard practice to split the IF amplifier into two parts, the first stage or stages being situated close to the mixer and connected to it by a few inches only of screened cable. These stages are called the Pre-Amplifier or Head Amplifier (see Fig. 728). The remaining IF stages are located in the main amplifier unit. Matching transformers or other devices are used, normally at both ends of the cable connecting the head amplifier to the receiver unit, the additional noise introduced by placing them at this point in the amplifier chain being negligible.

The first stage in the head amplifier is critical from the point of view of signal/noise ratio (see Sec. 9). Earlier receivers used a pentode in this stage, but more recent developments, particularly the production of crystals with lower noise temperatures, have enabled a substantial improvement to be obtained by the use of grounded-grid or neutralised triodes in head amplifiers.

35. The IF Amplifier

The gain of this amplifier should be sufficient to give full output from the detector on inherent noise alone, and this gain is easily calculated.

Consider the case of an IF amplifier with a pentode first stage following a crystal mixer; the mean square noise voltage $\overline{v^2}$ at the grid of this valve is given by equation (11) of Sec. 29.

$$\overline{v^2} = K \frac{(x \cdot t_{Gy} + G_d + 4.8 G_t)}{G^2} + R R_p \frac{(t_{Gx} + G_d + G_t)^2}{G^2} + R R_p \quad (5)$$

where x is the noise temperature ratio and t_{Gy} the transferred IF output conductance of the crystal, the remaining quantities being as defined in Secs. 10-12. Substituting the appropriate values for the CV1091 from Sec. 12 for an IF of 45 Mc/s and a crystal noise temperature ratio of 1.5, the equivalent RMS noise voltage at the grid is found to be 15.6 microvolts. A voltage amplification of the order of 10^6 will thus be needed to give full detector output. It is therefore usual to find about 5 stages in the IF amplifier proper, which, together with the two stages in the head amplifier, give a total of 7 IF stages.

The above argument applies only to receivers for centimetre wavelengths. At lower frequencies part of the required amplification is obtained at the signal frequency and the number of IF stages may then be reduced to 4 or 5.

The amplifier itself calls for little comment but it should be remembered that the necessity for providing an overall voltage amplification of about 10^6 at a frequency as high as 45 Mc/s involves very careful screening and the use of elaborate decoupling and filtering circuits in the supply leads. The required bandwidth is obtained either by staggered tuning or coupled tuned circuits or by a combination of the two (see Chap. 7).

36. Manual gain control

The gain of the receiver is varied by changing the gain of the IF amplifier by one of two usual methods:

- (1) The voltage applied to the screens is kept constant and a variable negative voltage applied to the control grids.
- (2) The control grids are kept at fixed potentials of about -1.5 volts and the screen voltages are varied.

The latter method is to be preferred since it is found that the change in input impedance of a RF amplifier valve is less marked when the mutual conductance is altered by varying the screen voltage than when the alteration is produced by a variation of the control grid potential. Since the combined screen currents of the controlled valves may exceed 10 mA the screens may be fed from the cathode of a Gain Control Valve, which is simply a cathode follower, the grid potential of which is determined by a potentiometer across the HT supply. By this means the use of a heavy duty potentiometer to supply the screens is avoided.

When strong signals are being received the gain control must be set near the position for minimum gain, and if the last IF stage were controlled it would be difficult to secure adequate output without overloading at the input. This stage is therefore operated with fixed electrode potentials. Gain control is not normally used on signal frequency amplifiers or IF head amplifiers as the operating conditions of these must be kept fixed if the optimum noise factor is to be maintained.

37. A.G.C.

Automatic gain control (AGC) is applied to communication receivers for the purpose of stabilising the carrier amplitude at the detector. This is accomplished by the standard method of rectifying the output from the last IF stage, filtering out the modulation frequency components and applying the resultant steady voltage (which must be of negative polarity) to the control grids of some of the signal frequency and IF valves. This method is suitable only when a continuous carrier is being radiated and cannot be applied to a radar receiver where the received signal consists of very short pulses of RF energy.

In radar reception it is the pulse amplitude at the display that must be stabilised and it is also essential to use some method of pulse selection so that AGC is operative only with respect to one particular echo. Fig. 729 shows a typical A-type display with the transmitter pulse (A) and five target echoes (B to F). One method of determining range is to set the echo required, for example D, with its leading edge on the vertical cross-wire, the range being read from the setting of the X-shift potentiometer. To ensure consistent range measurement the amplitude of D on the screen must be stabilised by AGC, the resultant amplitude being completely independent of the number and amplitudes of other target echoes which may be present.

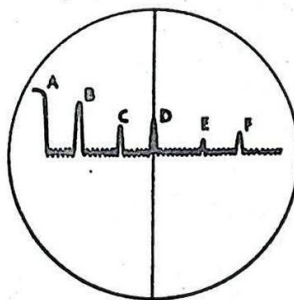
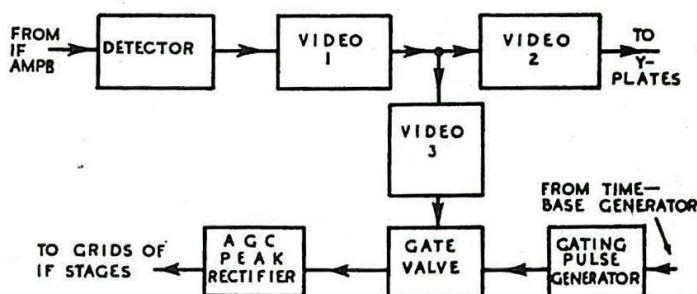


Fig. 729 -

A-type display showing transmitter pulse (A) and five target echoes (B to F).

The AGC voltage is obtained from a peak rectifier operated from the output of the video amplifier, a gating circuit being used to ensure that only the required echo is passed to the AGC rectifier. The circuit of a typical AGC system is shown in block diagram form in

Fig. 730; the gate valve circuit is given in Fig. 731 and the relevant waveforms in Fig. 732. There are two video stages between the detector and the cathode-ray tube, a third stage (VIDEO 3 of Fig. 730) being used to feed the AGC system. The output from the detector is of positive polarity



Block diagram of AGC system.

so that positive pulses are fed from the VIDEO 3 stage to the gate circuit. The output from the time-base generator is fed to the gate pulse generator which produces a short positive pulse as the spot passes across the centre line of the tube, the duration of this pulse being not much greater than that of a received echo. The pentode gate valve (Valve 2 Fig. 731) has its cathode held positive by virtue of the current through R_1 and R_2 and its current is normally cut off at both control and suppressor grids. The transmitter break-through pulse A and the received echoes B, C, E and F carry the control grid above cut-off, but no voltage is developed at the anode as the anode current is cut off at the suppressor grid. During the gating pulse the suppressor is at cathode potential and the pulse D applied to the control grid produces a negative-going pulse at the anode which is passed to the AGC peak rectifier. If no target echo is received during the gating pulse the sensitivity of the receiver rises until the inherent noise passes through the gate valve and operates the AGC rectifier, thus stabilising the height of the "grass" on the tube. Valve 1 of Fig. 731 is a clamping diode, preventing the suppressor from going positive with respect to the cathode (Chap. 12).

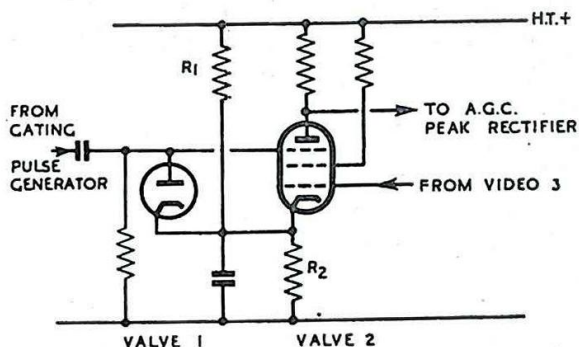
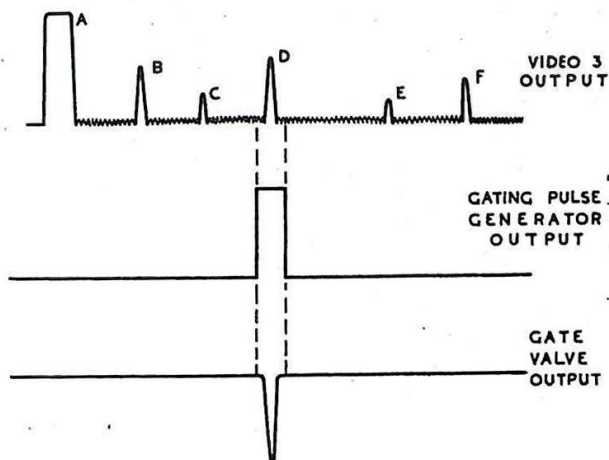


Fig. 731- Circuit of gate valve.

Fig. 732 - Waveforms of gating circuit.



38. Paralysis

The radar receiver is highly sensitive and is usually placed close to the powerful transmitter; in fact, with common T/R working the two are connected to the same aerial. A very large signal is thus received during the transmitter pulse and may result in a loss of receiver sensitivity for some little time after this pulse has ceased, the reception of echoes from nearby targets thus being rendered difficult or impossible. This temporary paralysis may occur in either grid or cathode circuits or in both, but it is convenient to consider the two cases separately.

(i) Grid paralysis

A positive voltage applied to the grid of any valve causes grid current to flow, and if the coupling is suitable, bias may be developed. In Fig. 733 the large voltage produced across the anode load of valve 1 during the transmitter pulse causes the grid of valve 2 to draw grid current, with the result that this grid is left negatively charged at the end of the transmitter pulse. This negative charge leaks away at a rate determined by the time constant $C_c R_g$ of the coupling circuit. Valve 2 is thus biased back at the beginning of the time-base sweep, or the current may even be cut off, so that the receiver sensitivity is reduced below its normal value and may not have recovered completely before the spot on the cathode ray tube has reached the end of its travel. The circuit of Fig. 734 is free from this defect, for although the condenser C_c is charged during the transmitter pulse the resistance of the tuned circuit to direct currents is so low that no appreciable bias voltage is developed across it. The charging and discharging of C_c now produce changes in the mean anode voltage of valve 1, which do not affect the receiver sensitivity to any marked extent.

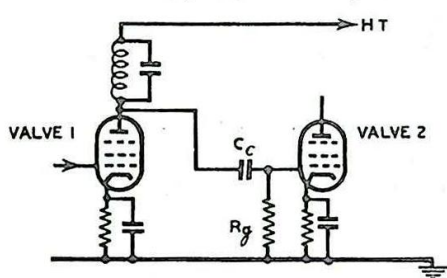


Fig. 733 -
Coupling circuit producing
grid paralysis.

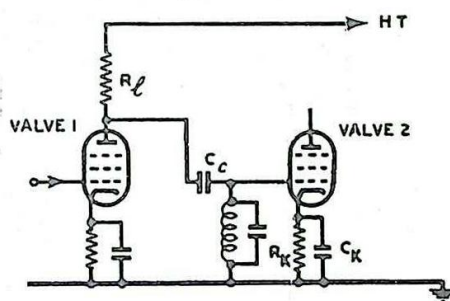


Fig. 734 -
Coupling circuit free from
grid paralysis.

The circuits of Figs. 733 and 734 show automatic bias for the amplifier stages. When fixed bias is used, in order to give manual gain control by variation of the screen voltage, as in Fig. 735, paralysis may be produced by the charging of the bias decoupling condenser C_b which must discharge through the decoupling resistor R_b before the grid can return to its normal potential. In this case paralysis may be

cured by connecting a large condenser, say 10 microfarads, across C_b . The resultant capacitance is not charged to any appreciable extent by the flow of grid current during the transmitter pulse, and while the mean rectified current must flow through R_b and thus produce a voltage drop across it, this is very small. C_b , which for 45 Mc/s. has a value of about 0.001 microfarad, must be retained, as electrolytic condensers are not effective for RF by-passing.

(ii) Cathode paralysis

During the transmitter pulse the grid receives a very large signal and rectification takes place in the anode circuit as well as the grid circuit, the mean cathode current rising well above its normal value. The cathode by-pass condenser, C_K of Fig. 734

then charges, and at the end of the transmitter pulse the cathode may be sufficiently positive for the valve current to be cut off. Even if this does not happen the sensitivity is reduced until the cathode has returned to its normal working potential.

Fig. 735 -

Coupling circuit for fixed bias.

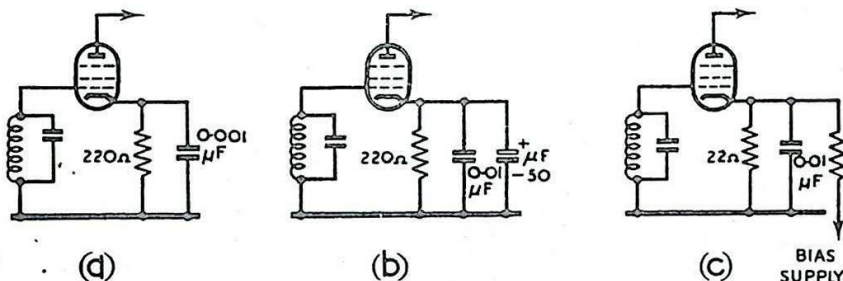
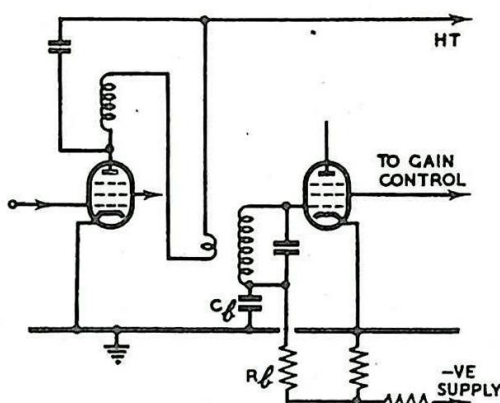


Fig. 736 - Circuits for eliminating cathode paralysis

Cathode paralysis may be dealt with in one of three ways:-

- (1) The time-constant of the cathode circuit may be made so short that at the end of the transmitter pulse the cathode returns very quickly to its normal potential. In Fig. 736 (a), which is suitable for an IF stage operating at 45 Mc/s., the by-pass condenser of 0.001 μ F is sufficient for thorough decoupling of the cathode at this frequency, while the time constant $C_K R_K$ is only 0.22 microsecond.
- (2) If the IF is considerably lower the by-pass condenser must be increased with a corresponding increase in the

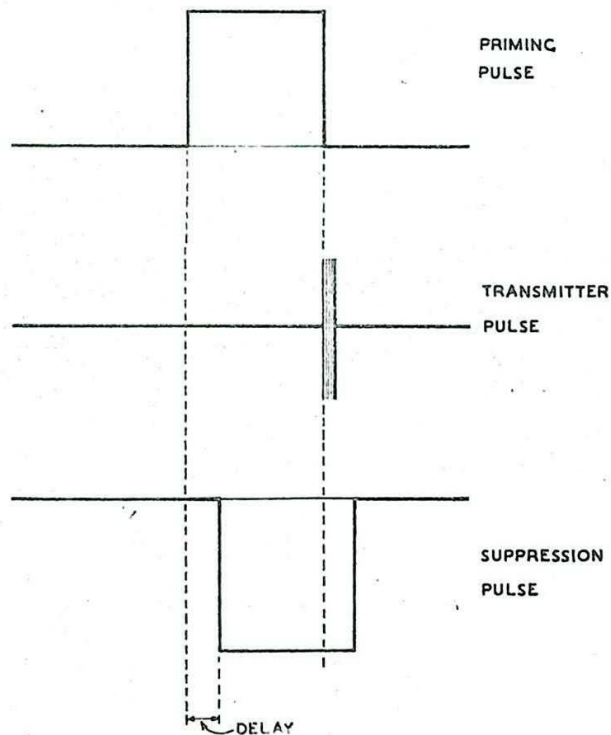
time constant, which may then become too long. In this case the RF by-pass condenser is shunted by a large electrolytic condenser (Fig. 736(b)), which prevents the cathode potential from rising to any appreciable extent during the transmitter pulse.

- (3) As an alternative to (2), the time constant may be kept low by decreasing the bias resistor, the additional current required to maintain the required bias being obtained from a separate bias supply (Fig. 736(c)). This system is rather wasteful in current, since the bias supply must deliver a considerably greater current than the normal cathode current of the valve (nine times as much in the case of the circuit of Fig. 736 (c)).

39. Receiver Suppression

If the large signal produced by the transmitter pulse can be prevented from reaching the receiver paralysis can be avoided. At first sight, the simplest method is to short-circuit the input during the transmitter pulse but this presents some difficulties. A mechanical relay is not sufficiently rapid, and all electronic switches have appreciable voltage developed across them by a large signal. With common T/R working such electronic switching is, of course, adopted, but the breakthrough from the transmitter pulse is still able to cause paralysis.

The most satisfactory method of overcoming these difficulties is to "suppress" the receiver during the transmitter pulse. In many radar equipments a 20 microsecond Priming pulse is used, the trailing edge of which coincides with the start of the transmitter pulse (Fig. 737). This priming pulse is inverted to give negative polarity, passed through a suitable delay circuit and is then applied as a suppression pulse to the suppressor grids of some of the signal frequency or IF stages of the receiver, thus rendering the latter inoperative during the transmitter pulse.



40. Temporal Gain (Swept Gain)

Fig.737- Waveforms of simple suppression circuit.

The signals received from nearby objects are very much stronger than those received from objects which are more distant. If the gain of the receiver is increased so as to receive the latter, the former may be strong enough to saturate the receiver and may even produce paralysis (Fig. 738). It is therefore desirable to have some means of varying the gain automatically as the spot moves across the screen, the

gain being low at the beginning of the trace and reaching maximum as the spot moves to the end of its travel. Instead of the simple square-pulse suppression waveform of Fig. 737 a more complicated waveform such as that of Fig. 739 is applied to the suppressor grids to give automatic variation of gain with range. After the transmitter pulse has ceased the voltage on the suppressors does not immediately rise to zero as in Fig. 737 but rises instead to a value just above suppressor cut-off (A of Fig. 739). During the time-base sweep the voltage rises exponentially to zero, the gain thus increasing along the time-base sweep. There are three variables which have to be adjusted for best results :-

- (1) the suppressor voltage at A,
- (2) the time-constant of the exponential rise AB, and
- (3) the delay time.

Of these the first is usually preset, the last two being under the control of the operator. When the controls are correctly adjusted all echoes are received at about the same amplitude, the appearance of the screen being roughly as in Fig. 740. Fig. 739 - Waveform of suppression pulse providing temporal gain. It will be noticed that as the gain increases along the time-base sweep the noise level rises also.

With a PPI display the setting of the temporal gain control differs from that needed with an A-type display since, even with equal echoes, the painting of the PPI is more intense nearer the centre. By the use of the controls, the suppression waveform may be adjusted to give a reasonably uniform painting of targets, at all ranges.

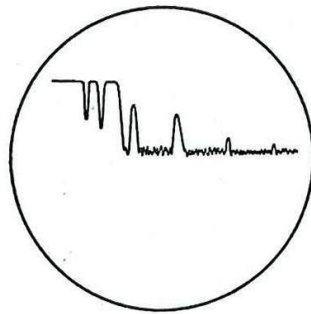


Fig. 738 - A-type display showing saturation produced by echoes from objects at close range.

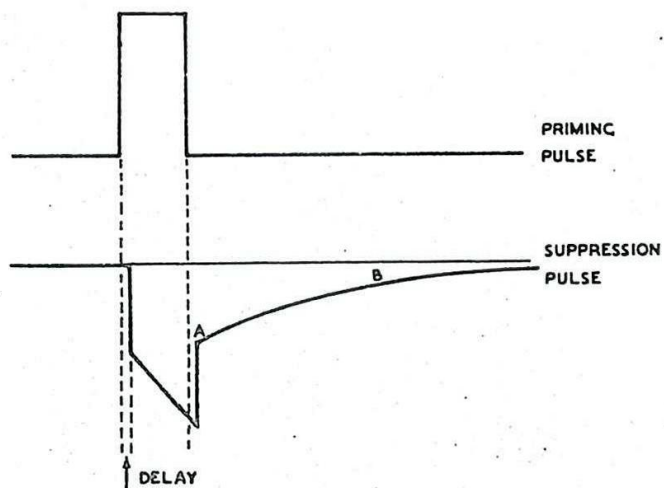


Fig. 739 - Waveform of suppression pulse providing temporal gain.

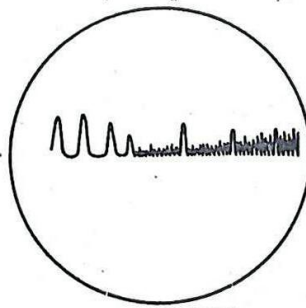


Fig. 740 - A-type display with temporal gain.

CHAPTER 17

AERIALS

1. FUNDAMENTAL CONSIDERATIONS

The direction-finding methods adopted in radar rely almost entirely on the use of an aerial system with known directional properties. Consequently, one is concerned here, much more than in communications, with the theory and design of aerial systems with specified directional performance. Again, in radar one is nearly always working with relatively weak received signals. Consequently, in the design of a radar set to give a specified range of detection, one requires to know the efficiency of the aerials in producing an electromagnetic field at the echoing object and in collecting or receiving back the small return signal or Echo. This aspect of aerials is referred to as the gain. The directional performance and gain are obviously closely connected; a highly beamed system will have a large gain. Finally, as in communications, one is often concerned with the impedance which the aerial presents to the transmission line connecting it with the transmitter or receiver. Often, for example, one requires an aerial whose impedance does not vary very much with change of frequency.

There are, therefore, three points in which one is usually interested when discussing an aerial: its directional properties, gain and input impedance.

A quantitative discussion of the working of the RF side of a radar system is best done from the point of view of power relationships. One is not so much concerned therefore with the number of amperes flowing in the transmitting aerial as with the number of watts it is radiating. Indeed, the form of the aerials (mirror, horn, waveguide, etc) usually makes it difficult to decide the point where an "aerial ammeter" might be connected. Also with receiving aerials one is not usually concerned with the "effective height" or received field strength or voltage, but rather with the "effective area" relating to the amount of power which the receiving aerial intercepts from the incident wave. Hence in what follows considerable emphasis will be laid on power relationships, and a minimum of discussion will be given on the flow of RF current in the aerials.

TRANSMITTING AERIALS

2. Type of Wave Emitted by a Transmitting Aerial

The radiated electromagnetic field, when observed in free space at a large distance from any aerial, is similar to the plane wave discussed in Chap. 5 Sec. 2. There is a transverse electric field \vec{E} and also a transverse magnetic field \vec{H} which lie in the wave-front, at right angles to one another and oscillating in phase. The wave-front travels out with velocity c , where

$$c = 3 \times 10^8 \text{ metres/second} \dots\dots\dots (1)$$

The wavelength λ is related to the frequency f of the transmitter by the relation

$$c = \lambda f \dots\dots\dots (2)$$

The wave-front is, however, only approximately plane. In fact it is spherical, with the aerial at the centre of the sphere. At large distances a portion of the sphere can be regarded as plane, but due to this spherical spreading the field becomes weaker as the distance r from the aerial increases.

This spherical spreading of energy is common to many branches of physics. The intensity depends on the surface area, the product of intensity and area remaining constant. For a sphere the surface area is proportional to r^2 . Thus the law for such phenomena is termed the Inverse Square Law. Hence the power P in an electromagnetic wave is given by :-

$$P \propto \frac{1}{r^2} \dots\dots\dots (3)$$

However, the electric field E is proportional to \sqrt{P} (see Sec. 7). Thus the electric field strength E follows the law :-

$$E \propto \frac{1}{r} \dots\dots\dots (4)$$

In what follows the electric field in a wave will usually be considered, but it must be remembered that the magnetic field also is necessarily present.

3. Field-Strength Diagram of an Aerial

The term Polar Diagram is commonly used synonymously with Field-Strength Diagram but strictly should be reserved for Field-Strength Diagrams plotted to Polar Co-ordinates.

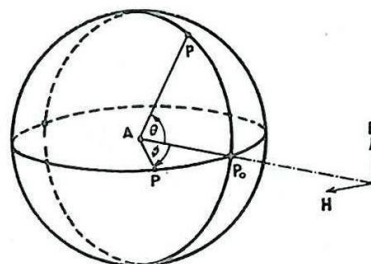


Fig.741.- Method of taking field strength diagrams.

In Sec. 2 the variation of field with distance was discussed, leading to equation (4). This formula applies when one travels out from the aerial in a fixed direction. It is desired now to consider how E varies when the point of observation is maintained at a fixed distance from the aerial but moves so that the line from the aerial to the observer varies in direction. This means that the point of observation P must be moved over a large sphere of radius r with the aerial A as centre (Fig.741). The value of E is required at all points on the sphere. In practice one often simplifies this by taking two perpendicular great circles, as shown in Fig.741, and observing first round one circle (θ variation) and then round the other (ϕ variation).

As remarked in Sec. 2, the radiated wave consists of \vec{E} and \vec{H} at right angles. They determine the Polarisation of the wave, e.g., horizontal polarisation if \vec{E} is horizontal and vertical polarisation if \vec{E} is vertical. We can therefore choose the planes of one of the great circles in Fig.741 (say the plane of θ -variation) to be parallel to the direction of \vec{E} and the other (plane of ϕ -variation) to be parallel to the direction of \vec{H} . The field strength diagrams are then said to be in the E -plane and H plane respectively.

There is nearly always one particular position P_0 on the sphere where the field strength is a maximum, since most aerials introduce a certain amount of beaming. It is convenient to make AP_0 (Fig.741) the line of intersection of the great circles and to measure the angles θ and ϕ from this direction.

If the field at P_0 is E_0 , then E at any other point P on the sphere will be expressible in the form

$$E = E_0 f(\theta, \phi) \dots\dots\dots (5)$$

$f(\theta, \phi)$ is the ratio of the electric field at any point (θ, ϕ) compared with the electric field in the direction of maximum field strength. The solid figure defined by $r = f(\theta, \phi)$ is termed the three-dimensional polar diagram. In the planes $\phi = 0$, $\theta = 0$, $r = f(\theta, \phi)$ gives two-dimensional polar diagrams in the E and H planes respectively. These are easily shown graphically and are indicative of the three-dimensional figure.

4. Field-Strength Diagram of Half-Wave Aerial

The half-wave aerial consists of a wire half a wavelength long with a gap at the centre to which a balanced feeder is attached (Fig. 742) leading from a transmitter. The current set up in the wire is not of uniform amplitude, being zero at the ends and a maximum in the middle. By analogy with standing waves on transmission lines or lecher bars, the current distribution expected is sinusoidal. However, owing to feeding and end-effects the distribution is only approximately sinusoidal in amplitude and the phase is not uniform. The exact results lead to difficult mathematical manipulation and the sine-wave amplitude distribution and uniform phase give results which are comparable with those obtained in practice. It will be assumed then, that the current has the same phase all along the aerial, the currents at all points on the wire rising and falling together but to different amplitudes. This is illustrated in Fig. 743. The

current in the wire arises from the movement of electric charges (e.g. electrons). In any small part of the wire, such as the point P in Fig. 743, the charges are first moving upwards then they are slowed down to rest. Next they are moving downwards and then brought to rest again. Consider a charge moving upwards with uniform speed (Fig. 744); lines of electric force will run out from this charge in all directions. Wherever the electric field is changing it is accompanied by a magnetic field the strength of which is proportional to the rate of change of the electric field. Now let the charge be brought to rest, i.e., decelerated. The electric lines will tend to go on but will be pulled back by the charge. "Kinks" will develop in the lines of force and will travel outwards, straightening out the lines of force to their correct positions relative to the charge (Fig. 745). We may regard the lines of force as being in tension, with the tension greater in the kink than in the line itself. The kink is the seat of an intense field

and is travelling outwards to form the electromagnetic wave radiated by the charge as its velocity is reduced to zero. The radial electric field is weak and negligible at large distances, but the kink is strong. The strongest stretching occurs in the kink of a line such as CD (Fig. 745) at right angles to AB, the direction of deceleration. In any other direction CF, making an angle θ with CD, the kink is less pronounced. It is the component of the kink at right angles to the electric line which constitutes the radiated field. The radiated field is therefore proportional to $\cos \theta$, this being the component of the deceleration of the charge at

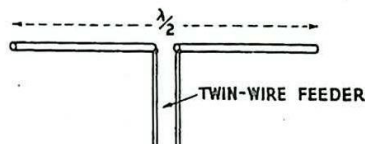


Fig. 742. - Half-wave centred aerial.

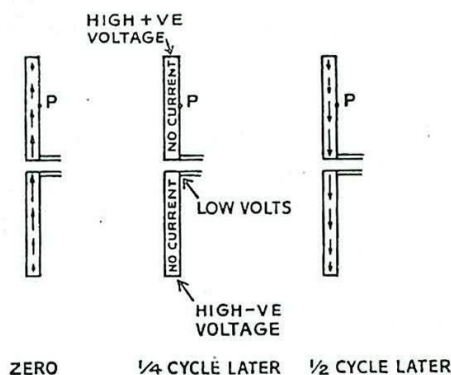


Fig. 743. - Current and voltage on half-wave aerial.

right angles to the electric line CF. Note that when the movement of the charge is sinusoidal instead of in jerks as hitherto considered the field is a maximum when the charge is at rest, i.e., when the current is zero. Thus the radiated field lags 90° on the current.

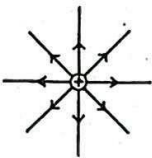


Fig.744.- Electric field around charge moving with uniform speed.

The charges all along the wire are radiating and the field is obtained by adding up the wavelets sent out by each accelerated (or decelerated) charge. Thus the following results may be deduced :-

- (i) Since the polarisation of the wave radiated from each charge is such that E always lies in a plane through the aerial, the same condition holds for the resultant wave radiated from the whole of the half-wave dipole.
- (ii) Since the dipole has axial symmetry, the field-strength diagram is symmetrical about a line drawn through the axis of the dipole; in other words, the H-plane field-strength diagram is given by

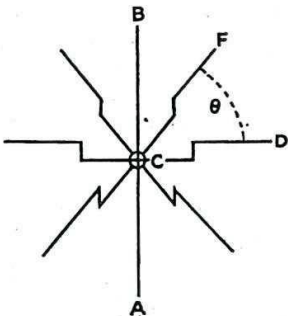


Fig.745.- Kinks in electric lines due to a deceleration of moving charge.

$$E = E_0 \dots\dots\dots (6)$$

In polar co-ordinates this equation represents a circle.

- (iii) The radiated field strength for each individual charge is proportional to $\cos \theta$, so that

$$E = E_0 \cos \theta \dots\dots\dots (7)$$

A similar result holds for the dipole as a whole. This result would be identical with (7) if we could assume that the wavelets from different charges along the aerial add together arithmetically. In fact, this is not exactly true. The wavelets from P_1, P_2 in Fig.747 do not arrive at a distant point P quite at the same time as a wavelet from the centre C of the wire. The wavelet from P_1 arrives earlier due to the path difference CD and that from P_2 arrives the same amount later due to the path difference P_2F . When we add together the effects from P_1 and P_2 we must therefore make a vector addition so as to

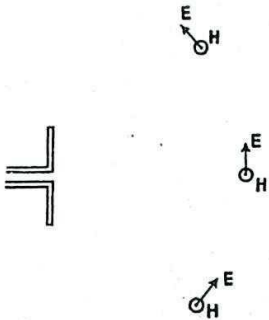


Fig.746.- Polarisation of wave from half wave aerial - E in plane of paper and H perpendicular to plane of paper.

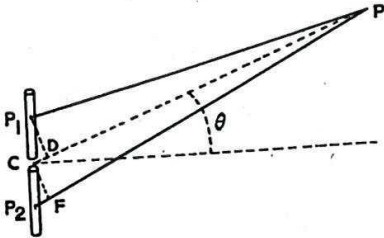


Fig. 747.- Phase difference between rays from pairs of points on half-wave aerial.

allow for the phase lead and lag (Fig.748). The resultant has the same phase as the wavelet from the mid-point of the aerial so that we may say, first of all, that the field appears to come from the mid-point C of the aerial. This point is called the Phase Centre. Next we note

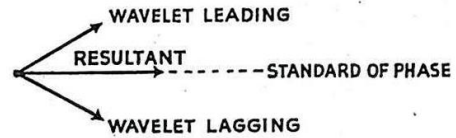


Fig.748.- Addition of wavelets.

that the phase lead or lag is greater as we move P_1 and P_2 out to the ends of the aerial. Failure to use the vector method of addition is therefore more serious for wavelets from near the ends of the aerial than for wavelets near the centre. The latter are, however, more important than the former since the current is strongest near the centre. The phase difference between wavelets gets larger as one moves round from $\theta = 0$ to $\theta = 90^\circ$. The wave is therefore rather less intense away from the $\theta = 0$ position than one would find simply by arithmetical addition of wavelets.

If the effects due to each elemental length are integrated over the length of the aerial the resultant field-strength diagram for a $\lambda/2$ aerial is given by :-

$$E = E_0 \frac{\cos \left[\frac{\pi}{2} \sin \theta \right]}{\cos \theta} \dots\dots\dots (8)$$

It is convenient to denote the E-plane Field-Strength Factor of a $\lambda/2$ aerial by a symbol, say D,

$$\text{where } D = \frac{E}{E_0} = \frac{\cos \left(\frac{\pi}{2} \cdot \sin \theta \right)}{\cos \theta}$$

.....(9)

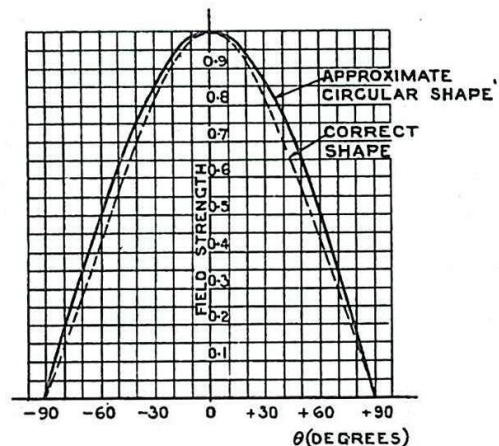


Fig.749.- Field strength diagram of half-wave aerial.

A plot of equation (7) is shown in Fig.749 with the curve corresponding to the correct equation (8) indicated dotted. It will be seen that the difference between the results for equations (7) and (8) is small. It is frequently convenient for mathematical derivations to assume that $D = \cos \theta$, the order of error being indicated. A plot in polar co-ordinates is often used, the radius from an origin 0 being taken to represent E at the appropriate angle. This method of plotting is shown in Figs.750 and 751 for equations (8) and (6) respectively. The field at a distance r measured out in any direction, not necessarily in the E - or H - plane, is similarly indicated by drawing radii of lengths proportional to E. The ends of these radii lie on a surface as shown in Fig.752.

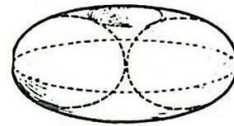
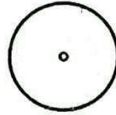
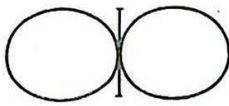


Fig. 750.- Field-strength diagram of half-wave aerial (E-plane); aerial in plane of paper.

Fig. 751.- Field-strength diagram of half-wave aerial (H-plane); aerial at right angles to plane of paper.

Fig. 752.- Field-strength diagram of half-wave aerial (all directions).

5. Isotropic Radiator

Although the half-wave aerial is the simplest met with in practice, its field-strength diagram as shown in Fig. 752 is still rather complicated and as a standard of reference it is convenient to consider a hypothetical aerial which radiates equally in all directions. Its field-strength diagram plotted in polar co-ordinates is a sphere. This aerial is sometimes called an Isotropic Radiator.

6. Energy Density in an Electromagnetic Wave

The electromagnetic wave propagated from an aerial is continually carrying away energy, which must be supplied by the transmitter. Recalling again that the wave consists of an electric field and a magnetic field, we see that energy is divided into two parts, electric energy carried by the electric lines of force and magnetic energy carried by the magnetic lines. These two energies are equally important and are in fact equal in magnitude in a plane wave. Electric energy by itself is familiar from consideration of a condenser charged up to a steady potential. Consider the condenser to be made of two plates each being one metre square, and placed one metre apart, with air dielectric; (Fig. 753).

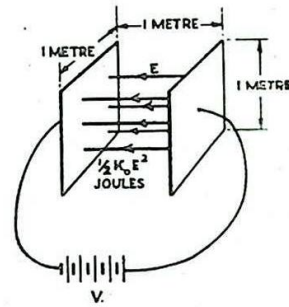


Fig. 753.- Electric field energy.

Its capacity K_0 is readily calculable to be

$$K_0 = \frac{10^{-9}}{36\pi} \text{ farads} \dots\dots\dots (10)$$

The energy, W (electric) stored in the condenser when there is a potential difference of V volts across the plates is given by the formula

$$W(\text{electric}) = \frac{1}{2} K_0 V^2 \text{ joules} \dots\dots\dots (11)$$

Regarding the electric field E in the space between the plates as the seat of this energy we may rewrite (11) as follows :-

Since the distance between the plates is 1 metre, a potential difference V volts means a field E volts/metre where

$$E = V \dots\dots\dots (12)$$

Also the volume of the space between the plates is 1 cubic metre. Hence (11) becomes :-

$$W_V(\text{electric}) = \frac{1}{2} K_0 E^2 \text{ joules/cubic metre} \dots\dots\dots (13)$$

and we can apply (13) to give the energy density of any steady electric field in air. Now in the wave, one is dealing with an alternating electric field. The same formula (13) will apply, however, provided E is taken to be the RMS value of the field strength. It has already been remarked that in a plane electromagnetic wave in space the magnetic energy is equal to the electrical energy. Thus the total energy density W_v for a wave whose electric field is E RMS volts per metre will be twice expression (13); i.e.,

$$\begin{aligned} W_v &= K_0 E^2 \\ &= \frac{10^{-9} E^2}{36\pi} \text{ joules/cubic metre} \dots\dots\dots (14) \end{aligned}$$

7. Power Relations for a Plane Electromagnetic Wave

Consider a wave of cross-sectional area one square metre (Fig. 754). Let an observer O stand near A watching the cross-section there. The wave is supposed to be moving to the right with velocity c metres/second given by equation (1). Now measure back from A a distance AB where

$$AB = c \text{ metres} \dots\dots (15)$$

The energy density W_v of the wave is given by (14). Hence the energy in the portion AB is

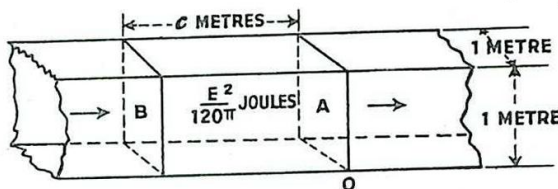


Fig. 754. - Power carried by a wave

$$c \cdot 1 \cdot 1 \cdot W_v \text{ joules;}$$

i.e., using (1) and (14)
the energy is

$$\frac{3 \times 10^8 \times 10^{-9} \times E^2}{36\pi} \text{ joules}$$

or, simplifying,

$$E^2/120\pi \text{ joules} \dots\dots\dots (16)$$

In one second the section BA will have moved to the right so that B is opposite the observer O , since B moves with speed c metres/second. Thus the energy that has passed through the cross-section at O in one second is that contained in the portion AB . This, with (16), leads to the important expression for the power P_a passing through a square metre of wavefront :-

$$\begin{aligned} P_a &= E^2/120\pi \text{ joules/sec/metre}^2 \\ &= E^2/120\pi \text{ watts/metre}^2 \dots\dots\dots (17) \end{aligned}$$

This equation is quite analogous to the formula for the power P dissipated in a resistance R ohms with an RMS voltage V across it :-

$$P = V^2/R \text{ watts} \dots\dots\dots (18)$$

From a comparison of (17) with (18) it follows that the factor 120π in (17) must have the dimensions of ohms. This number is therefore sometimes called the wave impedance, denoted by Z_w . We thus have

$$\begin{aligned} Z_w &= 120\pi \\ &= 377 \text{ ohms} \dots\dots\dots (19) \end{aligned}$$

and

$$P_a = E^2 / Z_w \dots\dots\dots (20)$$

The relations of this and the preceding paragraph apply to waves in air or free space. If the dielectric constant K and permeability μ of the medium are not unity the expressions have to be modified slightly by the introduction of K and μ into the equations.

8. Power and Field Relations for an Isotropic Radiator

The power from an isotropic radiator flows out equally in all directions. Thus if the transmitter puts P watts into the aerial it is possible to calculate the power passing through a square metre of wavefront at any distance. Suppose it is required to find the effect at distance r metres from the aerial. Draw a sphere of radius r metres, with the aerial as centre. The area A of the surface of the sphere is

$$A = 4\pi r^2 \dots\dots\dots (21)$$

The power P is passing out uniformly through this area. Hence P_a , the power per unit area of wavefront at distance r from this aerial, is given by

$$P_a = P / 4\pi r^2 \text{ watts/square metre} \dots\dots\dots (22)$$

Equations (17) and (22) enable the RMS field strength E to be expressed in terms of the input power P and the distance r . We have

$$\frac{E^2}{120\pi} = \frac{P}{4\pi r^2}$$

or

$$E = \frac{\sqrt{30P}}{r} \text{ RMS volts/metre} \dots\dots\dots (23)$$

for an isotropic radiator.

9. Power and Field Relations for Any Aerial

Even the simplest aerial, the half-wave aerial, does not transmit power equally in all directions, and in general, as remarked in Sec. 3, there is a "best" direction into which most power is sent. Suppose a given aerial is required to produce a field strength E_0 at a point distant r . The aerial is disposed so that the distant point lies in the aerial's best direction. It is then supplied with power P . This power will be less than the power, P' say, required by an isotropic radiator to give the same field at a distance r . The given aerial is thus more economical to use than the isotropic radiator and is said to have a Power Gain g where

$$g = P' / P \dots\dots\dots (24)$$

(As mentioned in Chap. 3 Sec. 5, gain is usually expressed as the logarithm of a ratio, whereas here it is simply the ratio.) The power gain of an aerial can be measured experimentally or, in a few cases, it may be calculated. If the power gain is known, then the performance of the aerial in its best direction is equal to that of an isotropic radiator, supplied with g times the power. Hence the field strength at distance r from an aerial supplied with power P is given by (23) after replacing P by gP , i.e.,

$$E_0 = \frac{\sqrt{30Pg}}{r} \text{ RMS volts/metre} \dots\dots\dots (25)$$

for an aerial of gain g , in its best direction.

Example

A highly beamed aerial used in a radar set has a power gain of 300. The power output is 100 kilowatts. What field strength would be detected by an enemy listening station 50 kilometres away ?

(a) From equation (25) we have

$$E_0 = \sqrt{\frac{30 \cdot 100,000 \cdot 300}{50,000}}$$

$$= 0.6 \text{ RMS volts/metre,}$$

(b) The answer may also be obtained from first principles, using the result that power per square metre is $E^2/120\pi$.

Thus :-

Suppose that an isotropic radiator is fed with 100 kilowatts. Consider a sphere of radius 50,000 metres. The area of the surface of this sphere is

$$4\pi \cdot 25 \cdot 10^8 \text{ square metres.}$$

Hence the power per square metre at the surface of this sphere is

$$\frac{10^5}{4\pi \cdot 25 \cdot 10^8} = 10^{-5}/\pi \text{ watts/square metre}$$

With a gain of 300 we should obtain

$$\frac{300 \times 10^{-5}}{\pi} \text{ watts/square metre.}$$

Hence required field strength E_0 is given by

$$\frac{E_0^2}{120\pi} = \frac{300 \times 10^{-5}}{\pi};$$

$$\text{i.e., } E_0^2 = 0.36$$

$$\text{or } E_0 = 0.6 \text{ RMS volts/metre.}$$

10. Power Gain of Half-Wave Aerial

The field-strength diagram of a half-wave aerial gives the variation of field strength E at a given distance r as the direction of the observation point is varied. Knowing the field-strength diagram of any aerial one can, in theory at least, deduce the gain in the following manner. Draw a sphere of radius r with the aerial at the centre and divide the surface up into small areas such as dA . Find from the field-strength diagram the field strength E at any small area on the sphere. The power flowing through the area is proportional to $E^2 \cdot dA$ and so a measure of the power coming from the transmitter is obtained by integrating $E^2 \cdot dA$ over the surface of the sphere. Now obtain from the field-

strength diagram the field strength E_0 in the best direction (and using the same units as before). From an isotropic radiator one would obtain E_0 all over the sphere's surface and the power required to maintain this would be proportional to $4\pi r^2 E_0^2$. The ratio of this to the power required by the given aerial gives the power gain. For the half-wave aerial we have approximately

$$E = E_0 \cos \theta$$

where E_0 can be taken as unity. Thus the power required is proportional to

$$\iint \cos^2 \theta \, dA$$

the integration being over a sphere of radius r . This gives $\frac{8\pi r^2}{3}$ and,

dividing into $4\pi r^2$, we obtain the result that the gain of a half-wave aerial is $3/2$ (1.8 db.) The gain is in fact a little greater than 1.5 since the field-strength is rather sharper than $\cos \theta$. An integration using the exact formula gives the correct result as 1.6 (or 2.04 db).

In general, for any aerial, we may write

$$E = f(\theta, \phi) \dots\dots\dots (26)$$

where $f(\theta, \phi) = 1$ in the best direction. Then

$$g = \frac{4\pi r^2}{\iint f^2(\theta, \phi) \, dA} \dots\dots\dots (27)$$

the integration being over a sphere of radius r . Equation (27) can also be written

$$g = \frac{4\pi}{\iint f^2(\theta, \phi) \, d\omega} \dots\dots\dots (28)$$

where $d\omega$ is the element of the solid angle $= dA/r^2$.

11. Input Impedance

Since the radiated wave carries power away from the transmitter an aerial can often be simulated by an impedance $R_r + jX_r$. The resistive part R_r of the dummy aerial must dissipate as heat the same amount of power as the aerial radiates. It is usually called the Radiation Resistance. The reactive part X_r corresponds to the storage of field energy in the vicinity of the aerial in the same way as energy is stored in a condenser or coil. Such stored energy near the aerial does not travel out into space like the radiated energy. Unlike the energy in a plane wave propagated out into space, the energy in the reactive storage field is not necessarily equally divided into electric energy and magnetic energy, and the aerial reactance may be capacitive or inductive according to which type of stored energy predominates. By special design, however, the two may be made equal and then the reactive term becomes zero. An aerial which is designed so that it has no reactive component of input impedance is termed a resonant aerial. The factors determining the resonant length are discussed later (Sec. 13). The shortest resonant length for a centre-fed aerial is just less than $\lambda/2$. It is usual to talk of a $\lambda/2$ aerial when the resonant length aerial is meant. The value of the radiation resistance is calculated from the current at the terminals and the power radiated. This involves a fairly detailed investigation and the result for the centre-fed half-wave aerial in free space is 73 ohms.

Fig.743 shows that the centre-fed half-wave aerial is being driven at a place where the current is high and voltage low, and the radiation resistance would be expected to be comparatively low. This type of aerial is said to be Current-Fed. If, however the feeder is attached to the end of the aerial as in Fig.755 it is fed at a place where the voltage is high and the current low. This method of End-Feeding is therefore referred to as Voltage-Feed. The input resistance for a voltage-fed half-wave aerial is very high, its exact value depending on the thickness of the aerial wire or tube. A voltage-fed half-wave aerial shown connected as in Fig.755 unbalances the feeder. Voltage-fed aeri-

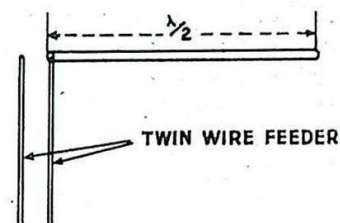


Fig.755.- Voltage feed to half-wave aerial.

are therefore often used in pairs, one on each side of the feeder, as shown in Fig.756. The pair is sometimes referred to as a Full-Wave Aerial or better, as an End-Fed Pair. The radiation resistance of such a pair of voltage-fed (or end-fed) half-wave aeri-

Radiation Resistance of a Pair of Voltage-Fed Half-Wave Aerials made of Cylindrical Tubing

(Fig.756)

Wavelength/diameter	Resistance in ohms
100	920
200	1300
300	1500
400	1700
600	2000
800	2200
1000	2400
2000	3000
3000	3300
4000	3600
6000	4000
8000	4300
10000	4600

The field-strength diagram and power gain of the pair is different from that of a single half-wave aerial and is discussed in Sec. 30.

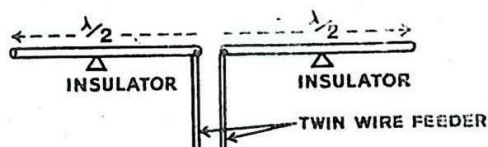


Fig.756.- Pair of half-waves; voltage fed (end-fed pair).

12. Q-Factor of a Resonant-Length Aerial

Only when it is cut to resonant length, i.e., a little less than $\lambda/2$, is the input impedance of a centre-fed aerial equal to a resistance of 73 ohms without reactance. When it is longer than the resonant length the impedance is inductive and when shorter it is capacitive. This may be deduced by comparing the half-wave aerial with a short section of open-circuited transmission line approximately $\frac{\lambda}{4}$ in length (Chap. 4 Sec. 14).

Thus if the aerial is cut to be a resonant length for a certain transmitter frequency f_0 and then the frequency is raised higher than f_0 the aerial becomes inductive, and if lower it becomes capacitive. The aerial thus behaves like a series tuned circuit, consisting as in Fig.757 of a coil L, condenser C and resistance 73 ohms.

The values of L and C are such that at frequency f_0 their reactances are equal and opposite so that they cancel out and the circuit appears like a resistance of 73 ohms. The input impedance (magnitude and phase) of the resonant length aerial is therefore of the form shown in Fig.758. One is usually interested in the selectivity or sharpness of the impedance curve, i.e., the Q -factor of the aerial.

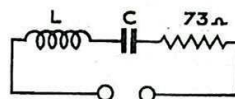


Fig.757.- Equivalent circuit of centre-fed half-wave aerial.

It is found that this depends on the thickness of the tubing or wire from which the aerial is made. Thin aerials have a higher Q than thick ones. The Q -factor of a resonant length aerial is given in the following table. This is based on theory, the results applying to either centre-fed or end-fed aerials. Few figures are available on the practical side but the values of Q obtained experimentally appear to be rather higher than the theoretical ones.

Theoretical Q -Factor of a Resonant length
Aerial made of Cylindrical Tubing

Wavelength/diameter	Q
10	1.6
25	2.7
50	3.5
100	4.3
200	5.2
300	5.7
400	6.0
600	6.5
800	6.8
1000	7.0
2000	7.9
3000	8.4
4000	8.7
6000	9.2
8000	9.5
10000	9.8

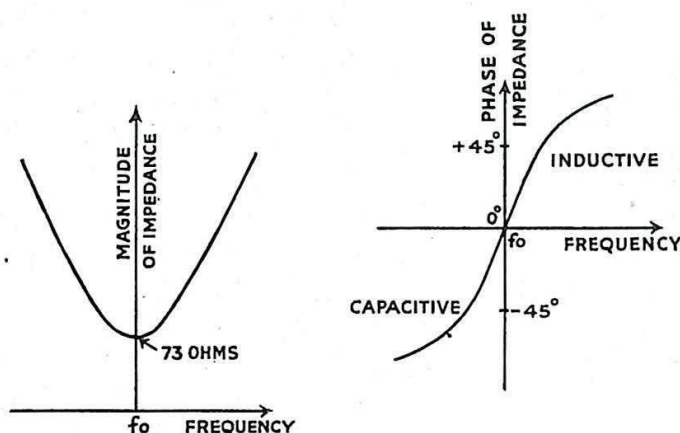


Fig.758.- Magnitude and phase of input impedance near resonance.

13. Factors Affecting the Resonant Length

As indicated previously, a centre-fed aerial in free space (e.g. suspended by fine strings from a high beam) is found to have a non-

reactive input impedance for a length rather less than 0.5λ . This appears to be due to the fact that the current distribution along the wire is not quite sinusoidal especially near the extremities. It is usually found that the non-reactive impedance occurs when the length is about 0.48λ . Curves of variation of R_r (for two values of the ratio wavelength to diameter) and X_r (for three values of the ratio) against length of aerial are given in Figs. 759 and 760, respectively. Note that the radiation resistance varies, but not so critically as the reactances. This variation in resistance probably accounts for the fact that the values of Q given in Sec. 12, and based on a constant resistance, are 10% to 20% too low.

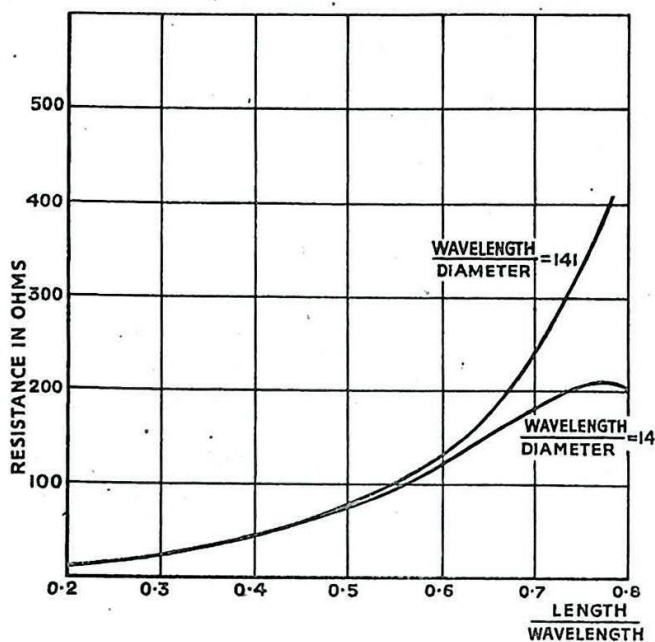


Fig. 759. - Resistance of centre-fed aerial.

In practice insulators have often to be used, and it is clear that an insulator fixed to an aerial at a place where the current is high and voltage small has little effect on the aerial. Thus end-fed resonant length aerials are usually supported at their centres (Fig. 756). An insulator fixed where the voltage is high will load the aerial like a condenser and make its effective length greater than its physical length. For this reason aerials are often cut shorter than the resonant length so as to allow for insulators at their extremities.

RECEIVING AERIALS

14. General

It is emphasised in Sec. 2 that the wave emitted from a transmitting aerial has a spherical wave-front but at large distances a small portion of the wave-front is effectively plane. Thus a receiving aerial is usually subject to the

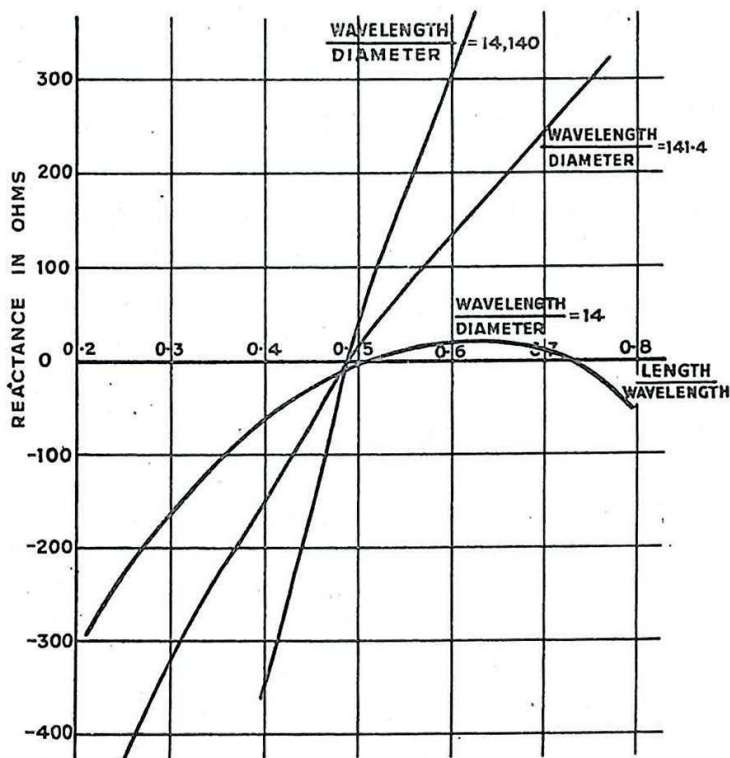


Fig. 760. - Reactance of centre-fed aerial.

incidence of a plane electromagnetic wave. This wave sets up an EMF in the receiving aerial. The EMF for a given incident field strength depends on the direction from which the wave is incident, and a plot of this EMF against direction gives the field-strength diagram of the receiving aerial.

In order to utilise the EMF v , a load z_l is attached to the terminals of the aerial. Current then flows in the load. However, this current also flows in the aerial, which consequently radiates some power away. Thus the current i in the load is not equal to $\frac{v}{z_l}$, but is given by

$$i = \frac{v}{z_l + z_a} \dots\dots\dots (29)$$

where z_a is the impedance of the aerial. The equivalent circuit of the receiving aerial corresponding to equation (29) is shown in Fig. 761. The useful power, i.e. the power in the load, is a maximum when the load is matched to the aerial impedance. This means, if the aerial has impedance R_a , that the load should be equal to R_a . If the aerial has a reactive component and impedance $R_a + jX_a$,

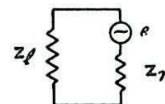


Fig. 761.- Equivalent circuit of receiving aerial.

then the load impedance should be equal to $R_a - jX_a$. (See Chap. 3 Sec. 4).

Under matched conditions, half the power is obtained in the load and half is re-radiated. In what follows it will always be assumed that the load is matched to the receiving aerial.

For a given incident field strength one is interested in how much power the aerial passes to the load. Owing to the non-uniformity of the field-strength diagram there is an optimum direction from which to have the wave incident. It is possible to compare matched receiving aeri- als by observing how much power is obtained in the load of each aerial when it is subject to a given incident field strength in its optimum direction. A hypothetical aerial which receives equally from all directions is taken as standard (Isotropic Receiver) and the ratio of the power picked up by the aerial to the power picked up by the isotropic receiver is the power gain of the aerial in reception.

Thus receiving aeri- als, like transmitting aeri- als, have a field-strength diagram, impedance and power gain. It has long been realised, almost intuitively, that these three properties of an aerial are the same when the aerial is used for reception and for transmission. A proof can be given, based on Maxwell's equations, and it is generally referred to as the Reciprocity Theorem.† In dealing with the properties of particular aeri- als one usually treats the aerial as a transmitter. The results are applicable also when the aerial is used for reception.

15. Power Relations for an Isotropic Receiving Aerial

The hypothetical isotropic receiving aerial is supposed to have a matched load attached to its terminals. If a plane wave of field-strength E RMS volts/metre is incident on the aerial, an EMF is set up in it and power is transferred to the load. As stated in Sec. 7, the power flow in the incident wave is (equation (17)),

$$P_a = \frac{E^2}{120\pi} \text{ watts/square metre} \dots\dots\dots (30)$$

Effectively, therefore, the aerial is intercepting some of this power and

† See Chap. 1 Sec. 7.

transferring it to the load. The aerial can therefore be represented by an area placed normal to the direction of the wave and of such a size that it intercepts the same amount of power as is found in the load. This is called the Effective Cross-Sectional Area of the aerial for absorption. Denoting it by A_0 square metres, the power P_0 in the load is given by :-

$$P_0 = P_a A_0 \text{ watts} \dots\dots\dots (31)$$

It can be shown that the effective cross-section of the isotropic absorber is given by the formula :-

$$A_0 = \lambda^2 / 4\pi \text{ square metres} \dots\dots\dots (32)$$

Thus the power in the load of an isotropic receiving aerial under the incidence of a field E is from (30), (31) and (32) :-

$$\begin{aligned} P_0 &= P_a A_0 \\ &= \frac{E^2}{120\pi} \cdot \frac{\lambda^2}{4\pi} \\ &= \frac{E^2}{120} \left(\frac{\lambda}{2\pi} \right)^2 \text{ watts} \dots\dots\dots (33) \end{aligned}$$

16. Power Relations for any Receiving Aerial

Even the simplest aerial, the half-wave aerial, does not receive equally from all directions. In its best direction it has a power gain of 1.6 over an isotropic aerial. In general, if an aerial has a power gain g , and if the wave is incident from the best direction, then the power in the load is g times the power in the load of an isotropic receiving aerial subject to the same wave. Alternatively one may say that the effective cross section A of the receiving aerial is g times that of the isotropic receiver A_0 . Hence we have

$$A = gA_0,$$

or, using (32),

$$A = \frac{g\lambda^2}{4\pi} \dots\dots\dots (34)$$

The power P in the load for an incident field strength E RMS volts/metre is given by

$$\begin{aligned} P &= P_a A \\ &= \frac{gE^2}{120} \left(\frac{\lambda}{2\pi} \right)^2 \text{ watts} \dots\dots\dots (35) \end{aligned}$$

Formula (34)(and also (32)) giving the effective cross-section for absorption by a receiving aerial is of great importance.

17. TRANSMISSION AND RECEPTION

The arguments of Secs. (9) and (16) enable one to solve the problem of calculating the power received from a distant transmitter Tx (Fig.762). Let the distance between transmitter and receiver be r metres. Let P_T be the transmitted power, g_T the gain of the transmitting aerial and g_R the gain of the receiving aerial. The power

passing through unit area of wavefront at distance r from the transmitter, in the best direction of the transmitting aerial, is

$$\frac{g_T P_T}{4 \pi r^2} \text{ watts/square metre} \dots\dots\dots (36)$$

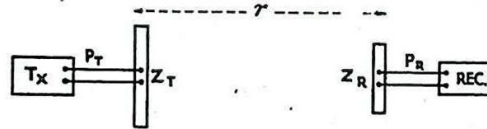


Fig.762. - Transmitter and receiver.

The effective absorption cross-section of the receiving aerial in its best direction is (see equation (34))

$$\frac{g_R \lambda^2}{4 \pi} \text{ square metres} \dots\dots\dots (37)$$

Hence the power P_R in the receiver is given by the product of (36) and (37) :-

$$P_R = P_T \frac{g_T \cdot g_R \lambda^2}{16 \pi^2 r^2} \text{ watts} \dots\dots\dots (38)$$

This power P_R is conveyed by feeder or waveguide to the receiver terminals (Fig.762). If the impedance looking into the receiver terminals is a resistance, R_i ohms, then the voltage V_R across the receiver terminals is given by

$$P_R = V_R^2 / R_i$$

$$\text{or } V_R = \sqrt{P_R R_i} \text{ RMS volts} \dots\dots\dots (39)$$

Example

A radio altimeter works on a frequency of 5,000 Mc/s., with a power output of $\frac{1}{2}$ watt. The transmitting aerial has a gain of 100 and is fitted under the aircraft and directed vertically downwards. If the aircraft is flying at a height of 3,000 metres find what signal would be detected by a man on the ground equipped with an aerial whose gain is 500.

Using formula (38), or from first principles, we have

$$P_R = \frac{P_T g_T g_R \lambda^2}{(4 \pi r)^2}$$

Here $P_T = \frac{1}{2}$, $g_T = 100$, $\lambda = 0.06$, $r = 3,000$.

Substituting, we find

$$P_R = \frac{(5 \cdot 10^{-1}) \cdot (10^2) \cdot (5 \cdot 10^2) \cdot (36 \cdot 10^{-4})}{(16 \pi^2) (9 \cdot 10^6)}$$

$$= \frac{25 \cdot 10^{-8}}{4} \text{ watts}$$

$$= 0.06 \text{ microwatts}$$

If the receiver input resistance is 80 ohms, what signal voltage would be obtained at the receiver ?

$$\text{We have } \frac{V^2}{80} = \frac{25 \cdot 10^{-8}}{4}$$

$$V^2 = 5 \cdot 10^{-6}$$

$$V = 2 \cdot 2 \cdot 10^{-3} \text{ RMS volts}$$

$$= 2 \cdot 2 \text{ RMS millivolts.}$$

PROPAGATION AND RECEPTION OF SHORT WAVES

18. Scattering by an obstacle

Consider a metallic obstacle, such as an aircraft, in the path of a radio wave. The obstacle acts as a receiving aerial and intercepts part of the power in the wave. A receiving aerial is usually matched to a load (Sec. 4), in which case half of the power intercepted goes to the load and half is re-radiated. When there is no load, as in the case of a perfectly-conducting metallic obstacle, all the power intercepted is re-radiated. The re-radiated power may be picked up by a receiving aerial placed near the original transmitter and this arrangement is then an example of radar.

The obstacle acts as a kind of relay transmitter between the actual transmitter and receiver. In practice most obstacles have complicated field-strength diagrams, but for simplicity it has become the custom to consider each aspect separately, e.g., for an aircraft, the tail-on, head-on and side-on aspects. Fixing on one aspect one notes the power received and assumes then that the obstacle is transmitting back to the receiver like an isotropic radiator. This simplifies calculations.

The result is that an obstacle, in a given orientation relative to the radar station, is assumed to intercept some fraction of the power incident on it and to re-radiate this power back to the radar station as if the obstacle were an isotropic radiator. If P is the power per square metre in the incident wave expressed in watts/square metre and P watts is the power re-radiated by the obstacle (considered as an isotropic radiator) then the effective area of the obstacle, its Echoing Area (or Equivalent Echoing Area) is A_e square metres where

$$P = P_a A_e \dots\dots\dots (40)$$

(compare equation (31)).

In radar, the wavelength is usually small compared with the dimensions of the scattering obstacle. This being so, the echoing area A_e of a flat metal plate of area A normal to the wave is given by :-

$$A_e = 4\pi (A/\lambda)^2 \dots\dots\dots (41)$$

It increases as λ , the wavelength, is decreased. On the other hand, most obstacles such as aircraft are curved. In this case the beneficial effect of reduction in wavelength is neutralised by the curvature becoming more important at small wavelengths. On the whole there is no simple relationship between A_e and λ . For a medium bomber A_e is about 1 square metre for centimetre wavelengths.

Example

A radar set works on a wavelength of 10 cm with peak pulse

power of 100 kilowatts. The same aerial is used for transmitting and receiving, its gain being 250. The minimum power detectable by the receiver is 10^{-12} watts. What is the maximum range of detection on a medium bomber?

The flux of power at Range R metres from the transmitter is given (equation (36)) by

$$\begin{aligned} P_a &= \frac{g_T P_T}{4\pi R^2} \\ &= \frac{250 \cdot 10^5}{4\pi R^2} \\ &= \frac{6 \cdot 25 \cdot 10^6}{\pi R^2} \text{ watts/square metre.} \end{aligned}$$

The echoing area is assumed to be one square metre (see Sec. 18). Hence the power P_r re-radiated is

$$P_r = \frac{6 \cdot 25 \cdot 10^6}{\pi R^2} \text{ watts (42)}$$

If this is radiated isotropically, the power flux from the aircraft at the receiving aerial is

$$\frac{P_r}{4\pi R^2} \text{ watts/square metre (43)}$$

The gain of the receiving aerial is 250, so its effective area is

$$\begin{aligned} &\frac{250 \lambda^2}{4\pi} \\ \text{i.e. } &\frac{6 \cdot 25}{10\pi} \text{ square metres (44)} \end{aligned}$$

The power received is the product of (43) and (44); i.e.,

$$\begin{aligned} &\frac{P_r}{4\pi R^2} \cdot \frac{6 \cdot 25}{10\pi} \\ &= \frac{6 \cdot 25 \cdot 10^6}{4\pi^2 R^4} \cdot \frac{6 \cdot 25}{10\pi} \\ &= \frac{3 \cdot 2 \cdot 10^4}{R^4} \text{ watts.} \end{aligned}$$

This is to be equal to the minimum detectable signal of 10^{-12} watts. Thus

$$\frac{3 \cdot 2 \cdot 10^4}{R^4} = 10^{-12}$$

so that $R^4 = 3 \cdot 2 \cdot 10^{16}$

and $R = 13000 \text{ metres} = 13 \text{ km} \approx 8 \text{ miles.}$

19. Range of a Radar Set

The above example illustrates the method of calculating the maximum Range of a radar set. The general formula is obtained as follows :-

Let P be the transmitted power in watts, λ the wavelength in metres, and g the gain of the aerial (used for transmitting and receiving). Then at range R metres the flux of power P_a is given by

$$P_a = \frac{gP}{4\pi R^2} \text{ watts/square metre} \dots\dots\dots (45)$$

(compare (22) and Sec. 9).

Now let the echoing area of the target be A_e square metres. Then the power P' re-radiated by the target is :-

$$\begin{aligned} P' &= P_a A_e \\ &= \frac{gP A_e}{4\pi R^2} \text{ watts} \dots\dots\dots (46) \end{aligned}$$

The flux P'_a of power, back at the receiver, distant R metres from the target is

$$\begin{aligned} P'_a &= \frac{P'}{4\pi R^2} \\ &= \frac{g^2 A_e}{(4\pi R^2)^2} \text{ watts/square metre} \dots\dots\dots (47) \end{aligned}$$

The effective cross-section A of the receiving aerial is given by (34), i.e.,

$$A = \frac{g\lambda^2}{4\pi} \text{ square metres.}$$

Thus the power P'' received is

$$\begin{aligned} P'' &= AP'_a \\ &= \frac{g^2 \lambda^2 P A_e}{(4\pi)^3 R^4} \text{ watts.} \end{aligned}$$

If the minimum detectable power is \hat{P} , then the maximum range of detection \hat{R} is given by the equation

$$\begin{aligned} \hat{P} &= \frac{A_e P \lambda^2 g^2}{(4\pi)^3 \hat{R}^4} \\ \text{or } \hat{R} &= A_e^{\frac{1}{4}} \cdot (P/\hat{P})^{\frac{1}{4}} \cdot \lambda^{\frac{1}{2}} \cdot g^{\frac{1}{2}} \cdot (4\pi)^{-\frac{3}{4}} \text{ metres} \dots\dots (48) \end{aligned}$$

The aerial is often a circular mirror of diameter $2r$ and area $\pi(2r)^2/4$. For such a mirror it can be shown (equation (76)) that

$$\begin{aligned} g &= \frac{4\pi}{\lambda^2} \cdot \frac{\pi(2r)^2}{4} \\ g^{\frac{1}{2}} &= \frac{2\pi r}{\lambda} \end{aligned}$$

Thus we find that

$$\hat{R} = A_e^{\frac{1}{4}} \cdot (P/\hat{P})^{\frac{1}{4}} \cdot 2r \cdot \lambda^{-\frac{1}{2}} \cdot (\pi/64)^{\frac{1}{4}} \text{ metres} \dots\dots (49)$$

It is remarkable that the maximum range increases only as the fourth root of the transmitted power. It is directly proportional to the diameter of the mirror.

20. Reflection from Ground and Sea

In most of the preceding sections it has been assumed that the aerial has been in free space, i.e., entirely independent of its position with respect to the ground or sea. It will be shown in the following sections that the field-strength diagram of any aerial can be considerably modified by reflections from the ground or sea. In order to investigate the behaviour of radio waves incident on the ground one must know the electrical properties of the earth. Consider a condenser made up from two metal plates with the space between filled with soil. The earth is not a very good insulator and the arrangement is most simply represented as a resistance or, better, a conductance G in shunt with a capacitance, C . When a high frequency alternating current is applied to the plates the admittance is given by

$$G + j \omega C \dots\dots\dots (50)$$

where $\omega/2\pi$ is the frequency. It follows from expression (50) that for low frequencies the term G will be bigger than the term C and thus the earth behaves like a metal at low frequencies. On the other hand at sufficiently high frequencies ωC will be bigger than G and the earth will behave as a dielectric.

The conductivity of the earth varies considerably depending on the dampness and type of the soil or rocks. However, when average values are substituted in (50) it appears that for all wavelengths used in radar, the ωC term predominates and thus the earth can be considered to be a dielectric. The dielectric constant is generally taken to be 10.

Now consider the case of the sea. Again one takes a condenser with sea water between the plates and arrives at expression (50) for the admittance. The conductivity of the sea is much higher than that of land and the G term is therefore much larger. The frequency at which the conductive and susceptive terms of (50) are equal is approximately 1000 Mc/s (wave length 30 cm.). For frequencies well below this value the sea behaves like a metal whilst for much higher frequencies it behaves like a dielectric. Broadly speaking, on metre wavelengths the sea acts as a metal and on centimetre wavelengths the sea acts like a dielectric. The dielectric constant is 81.

21. Horizontally Polarised Waves Incident on Ground

As explained above, the ground acts as a dielectric for radar wavelengths. The incidence of a wave on the ground is thus similar to light falling on a glass surface. Part of the wave is reflected and part refracted. For normal incidence the well-known optical law,

$$\text{fraction reflected} = \frac{n-1}{n+1}$$

applies, n being the refractive index. The index of refraction of glass is 1.5 and of the earth 3.2 (square root of dielectric constant). Consequently, reflection of radio waves from the surface of the earth is more pronounced than that of light waves from glass.

Suppose now that the wave is polarised with the electric vector horizontal. At the point of incidence of the ray on the earth, we have

three electric fields $\vec{E}(\text{In})$, $\vec{E}(\text{Refrac})$ and $\vec{E}(\text{Reflec})$, for the incident, reflected and refracted rays. The two just above the surface must add together to equal the one just below the surface (Fig.763), i.e.



Fig.763.- Horizontally polarised radar waves incident on earth.

$$\vec{E}(\text{In}) + \vec{E}(\text{Reflec}) = \vec{E}(\text{Refrac}) \dots\dots\dots (51)$$

Owing to the high index of refraction, most of the wave is reflected and $E(\text{Refrac})$ is small, so that we have approximately, from (51),

$$\vec{E}(\text{Reflec}) = -\vec{E}(\text{In}) \dots\dots\dots (52)$$

There is thus a sudden phase change of about 180° on reflection and the reflection coefficient is practically unity. Reflection is most complete when the rays strike the surface at a small glancing angle, and falls off as one approaches normal incidence. The fraction reflected at normal incidence is about 50%.

22. Horizontally Polarised Waves Incident on Sea

Take first the case of centimetre wavelengths. As explained in Sec. 20, the sea then acts like a dielectric, and similar considerations arise as in the previous paragraph. Reflection is, however, more complete owing to the very high index of refraction, viz. $\sqrt{81}$ or 9. Even at normal incidence the reflected ray is 90% the amplitude of the incident.

In the case of metre waves the sea behaves substantially metal. Take an incident and reflected wave as shown in Fig.764. The resultant electric field at the surface must be small (Chap. 5 Sec. 3). Hence the electric fields of the incident and reflected waves must be roughly equal and opposite. The reflection coefficient is about unity and there is a phase change of 180° on reflection.

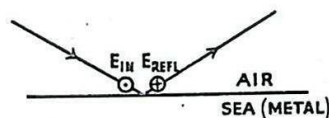


Fig.764.- Horizontally polarised metre waves incident on sea.

Thus, with horizontally polarised waves of any radar frequency, incident on the sea, the reflection coefficient can be assumed to be unity with a phase change of 180° . The sea is liable to be rough and, for short wavelengths, the roughness is of the same order as the radio wavelength. Reflection is not then specular and a definite reflected wave may not be found.

23. Reflection of Vertically Polarised Waves

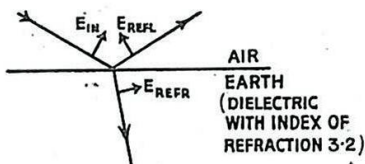


Fig.765.- Vertically polarised radar waves incident on earth.

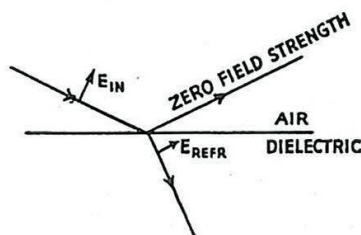


Fig.766.- Brewster angle.

The reflection of horizontally polarised waves discussed in Sec. 21 and 22 is comparatively simple, and the general result is that reflection is nearly always practically complete but with 180° phase change. When the electric field of the wave is vertical the situation is more complex. In this case the electric vectors in the incident, reflected and refracted waves are not parallel to the surface (Fig. 765) and the fitting together of the fields across the boundary is a comparatively difficult process.

Take first the case of any radar waves incident on the earth or of centimetre waves incident on the sea. The medium then behaves like a dielectric. At one particular angle of incidence the reflected and refracted rays are at right angles and the electric field E (Refrac) of the refracted wave is parallel to the direction of the reflected ray (Fig. 766). The electrons in the dielectric, oscillating under the influence of E (Refrac) radiate nothing along the direction of E (Reflec) and so the reflected wave is zero. This is the Brewster Effect, the appropriate angle of incidence being called the Brewster Angle. Measured from the horizontal it is about 17° for earth and 62° for sea. The Brewster effect will occur when the earth or sea is behaving like a dielectric, i.e., for all radar wavelengths incident on earth and for centimetre wavelengths incident on sea. At angles to the horizontal below the Brewster angle, the reflection takes place with phase reversal (Fig. 767); above the Brewster angle, without phase change (Fig. 768).

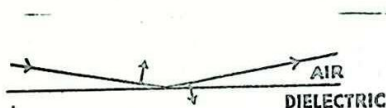


Fig. 767.- Reflection of vertically polarised waves below Brewster angle (Note: The refracted ray is not shown).

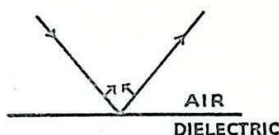


Fig. 768.- Reflection of vertically polarised waves above Brewster angle (Note: The refracted ray is not shown).

In general, subject to certain reservations discussed in Sec. 26, vertically polarised metre waves are reflected from the sea without phase change and with practically perfect reflection. The resultant field parallel to the surface is zero, the sea acting like a metal for these wavelengths.

24. Effect of Flat Earth on Field-Strength Diagram (Horizontal Polarisation)

Consider a horizontal half-wave aerial A placed height h above flat ground; (Fig. 769). Take a ray going off at an angle α above the horizontal and another at an angle α below the horizontal. These rays are equal in strength. The down-going ray hits the earth or sea and as indicated in Sec. 21 and 22 it may be assumed to suffer 100% reflection with 180° phase change if the angle of incidence is not near the perpendicular direction. At a far distant point P the direct ray AP and the indirect ray ABP come together and have to be added

vectorially. Draw AN perpendicular to the surface and produce to A' with $A'N = AN$ and join A'B. Then from the Fig. 769,

$$A'B = AB.$$

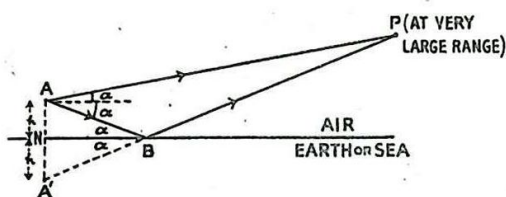


Fig. 769.- Horizontal half-wave aerial and image.

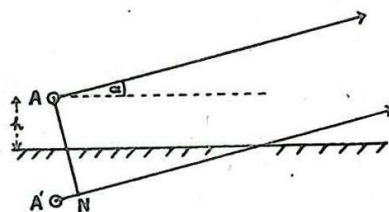


Fig. 770.- Isotropic radiator and image.

Thus, so far as the path length is concerned, the ray ABP can be replaced by the path A'BP. Indeed one can put an image source at A', consisting of another half-wave aerial fed with equal power and imagine the earth to be removed. Since the ray ABP suffers 180° phase change at B, the image source A' must oscillate 180° out of phase with A (Fig. 769).

Consider now an isotropic source a height h above a flat earth and radiating horizontally polarised waves. The image is a depth h below the earth. The field-strength diagram of the pair is found by taking a pair of parallel rays, from the source and its image, making an angle α with the horizontal. The fields are added vectorially at a distant point. Referring to Fig. 770, AN is drawn from the source A perpendicular to the lower ray. The path difference of the two parallel rays is A'N; in addition there is the fixed phase difference of 180° between source and image.

When the angle α is very small the path difference A'N is also small and the total phase difference is 180° so that the resultant is zero. Thus no wave is propagated in the direction of grazing incidence along the earth's surface. As the angle α is gradually increased from zero the total phase difference between the two rays increases from 180° towards 360° , at which the two rays are in phase and reinforce. This condition arises when A'N is $\lambda/2$. For a given λ , if the distance $2h$ between the source and image is very large, and α , need not increase greatly from zero in order to make $A'N = \lambda/2$. The bigger the base-line, so to speak, the sooner the path difference develops.

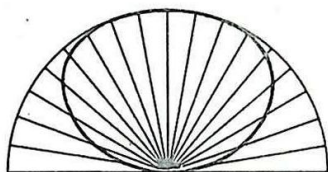


Fig. 771.- Interference factor for aerial distance $\lambda/4$ above earth.

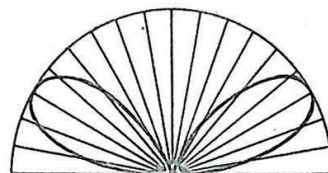


Fig. 772.- Interference factor for aerial distance $\lambda/2$ above earth.

To take a few examples, suppose $h = \lambda/4$, then α must increase to 90° , with the rays shooting straight up, before a total phase difference of 360° is achieved. The field-strength diagram in polar co-ordinates is then as shown in Fig. 771. When $h = \lambda/2$, the phase difference is $+180^\circ$ for rays shooting straight up, and the field strength diagram is as illustrated in Fig. 772. As the height is increased, more and more lobes appear, there being a lobe between 0° and 90° for every $\lambda/2$ of height above the ground. These are often called Interference Lobes. For an aerial system 7λ above the earth there will be 14 lobes as shown in Fig. 773. The lobes at low angles are approximately equally spaced in angle and occur at $2^\circ, 6^\circ, 10^\circ, 14^\circ \dots$ but at high angles the angular spacing opens up, with the lobes more widely separated. For 14λ height above the ground, the first lobes would be at $1^\circ, 3^\circ, 5^\circ, 7^\circ \dots$ and so on. The angular difference between the lowest lobes is inversely proportional to the height of the phase-centre of the aerial system.

When the beaming of the aerial system is not too strong, so that the first interference lobe occurs near the maximum of the free-space field-strength diagram, the field strength at a given distance in the direction of the first lobe is twice the free-space value. Thus by equation (25) the power gain g , of the system is four times the power gain of the aerial in free space and the maximum range of detection of a ground radar system is twice the free space value; (see equation (48)).

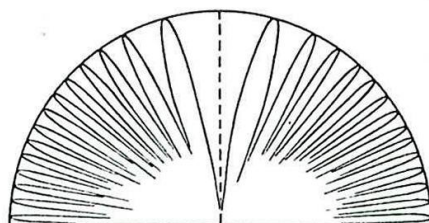


Fig. 773.- Interference factor for aerial distance 7λ above earth.

25. Effect of Flat Earth on Field Strength Diagram (Vertical Polarisation)

From the results of Sec. 24 and 23 we deduce the following facts for vertically polarised radar waves incident on ground. Up to the Brewster angle (17°), the effects are similar to those for horizontally polarised waves, with lobes depending on the height of the aerial system above the ground and as in Sec. 24. Near the Brewster angle the reflected ray disappears. Above the Brewster angle reflection takes place without phase change, and broadly speaking, maxima and minima in the field-strength diagram are interchanged from their positions in the horizontally polarised case (Fig. 774).

For vertically polarised centimetre waves incident on a smooth sea the same results hold but the Brewster angle is now 62° .

For vertically polarised waves of several metres in wavelength incident on the sea, there is no phase change on reflection. The maxima and minima of the horizontally polarised case are here interchanged and, in particular, there is a lobe at sea level (Fig. 775).

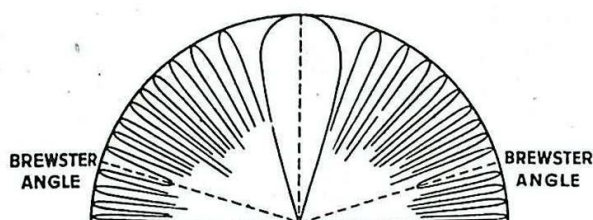


Fig. 774.- Interference factor (vertically polarised waves) for aerial distance 7λ above earth.

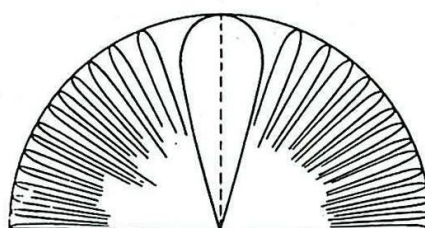


Fig. 775.- Interference factor (vertically polarised metre waves) for aerial distance 7λ above sea.

26. Summary of Effect of Flat Surface

The following table shows the lobes in the vertical field-strength diagram of an isotropic radiator placed 14λ above earth or sea and summaries the results of the two preceeding paragraphs :-

<u>Polarisation</u>	<u>Wavelength</u>	<u>Surface</u>	<u>Lobes</u>
Horizontal	Any radar	Land	$1^\circ, 3^\circ, 5^\circ$ and so on.
Horizontal	Any radar	Sea	$1^\circ, 3^\circ, 5^\circ$ and so on.
Vertical	Any radar	Land	$1^\circ, 3^\circ, 5^\circ$ up to 17° then change over to $18^\circ, 20^\circ, 22^\circ$ and so on.
Vertical	Centimetre	Sea	$1^\circ, 3^\circ, 5^\circ$ up to $6\frac{1}{2}^\circ$ then change over to $8^\circ, 10^\circ, 12^\circ$ and so on.
Vertical	Metre	Sea	$0^\circ, 2^\circ, 4^\circ, 6^\circ, 8^\circ$ and so on.

The free-space field-strength diagram is split up into lobes depending on the height of the phase centre above the earth's surface. Mathematically, the free-space field-strength factor must be multiplied by

$$\sin \left[\frac{(360^\circ h \sin \alpha)}{\lambda} \right] \dots\dots\dots (53)$$

or

$$\cos \left[\frac{(360^\circ h \sin \alpha)}{\lambda} \right] \dots\dots\dots (54)$$

according to whether there is or is not change of phase at reflection. Although the maximum range of detection is increased by the presence of the lobes, the zeros in between the lobes give gaps in the radar coverage.

Generally speaking, reflection is more complete and the formation of maxima and minima is more straightforward with horizontally polarised than with vertically polarised waves. With the metre waves used in radar, one is inevitably working over the wavelength band in which the sea is changing from a "metal" to a "dielectric". This changeover does not take place suddenly and so the simple treatment given above shows only the essential details of the effect of the sea on vertically polarised waves.

To take an example, suppose we are dealing with 3-metre waves, vertically polarised, and incident on the sea. A detailed investigation shows that there is a pseudo-Brewster angle at 2° . The lobe at sea level is thus unlikely to be well developed. For wavelengths of a metre and upwards this pseudo-Brewster angle is inversely proportional to the square root of the wavelength - at 30 metres it would be about $2/3^\circ$ - and only on quite long wavelengths will the sea give exactly the effects of a good conductor at very low angles of elevation.

The choice of polarisation for a coastal or sea-borne radar using metre waves to detect low-flying aircraft has been a matter of some controversy. Observations show that with vertical polarisation there is considerable sea-clutter (i.e., scattering back from sea-waves) up to ranges of about 20 miles, but with horizontal polarisation sea clutter is small. On the other hand the ranges of detection of low-flying aircraft are about the same for either polarisation. It appears therefore that there is more field strength at sea level with vertically polarised waves, but a little above the sea the field strength of horizontally polarised waves soon becomes appreciable. The results obtained clearly depend on the height of the aerials above the sea, and the matter is thus further complicated. Most British radar sets for aircraft and ship detection use horizontally polarised waves.

27. Gap Filling

One method of filling the gaps in the vertical field-strength diagram is to employ two aerial systems at different heights above the ground. These can be arranged so that many of the gaps in one field-strength diagram are filled by the lobes of the other and, by switching as required, a target may be followed continuously.

Another method is to use a combined array of vertically and horizontally polarised aerials. From the previous paragraphs the maxima and zeros are interchanged when the polarisation is changed. The resultant field-strength diagram is gapless but not necessarily uniform. Care must be taken with the feeding to give the correct phasing to ensure that the signals from the two arrays are additive.

28. Effect of Earth's Curvature : Refraction

Consider an aerial system A at height h feet above the earth's surface. The earth is approximately a sphere of radius 4,000 miles and the geometric horizon for this aerial system is at the point P in Fig. 776. Although radio waves of large wavelength are diffracted easily round the earth's curvature, short waves as used in radar do not diffract

appreciably into regions beyond and underneath the optical horizon.

Strictly speaking the horizon is slightly more distant for radio waves than the geometric point in Fig. 776. This is due to the fact that these waves are bent or refracted in the earth's atmosphere.

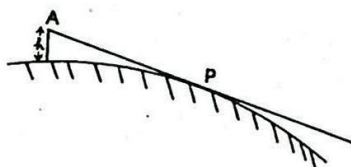


Fig. 776. - Optical horizon.

Under normal conditions, the dielectric constant of air is 1.00055. Higher up in the atmosphere the air becomes less dense and the dielectric constant approaches more nearly to unity. Thus a ray travelling up into the atmosphere is continually passing from more dense to less dense regions in the optical sense. In accordance with the principles of refraction, the ray is therefore continually bent away from the normal. The effect is very small but is made more appreciable by the presence of water vapour in the atmosphere. Owing to its molecular structure, water vapour has a dielectric constant of 1.01 at normal temperature. Thus, although the amount of water vapour in the atmosphere is only 1.4%, its efficacy as a refracting medium is comparable with that of the air itself. The percentage of water vapour decreases as one ascends into the atmosphere so that again we have the effect of bending the radio rays. A ray starting out at a small angle to the horizontal is therefore slowly bent round and can reach a point more distant than the geometrical horizon. It is found that this can be allowed for approximately by assuming the radius of the earth to be 6,000 miles instead of 4,000.

The effect just described is normal refraction. Occasionally in tropical climates the water vapour content is high, and varies very rapidly with height. In this case Super-Refraction takes place, the rays may even bounce several times from the earth's surface, and considerable penetration takes place into the region beyond and beneath the geometrical horizon. This phenomenon is also termed Anomalous Propagation or briefly Anoprop.

COMMON TYPES OF AERIAL ARRAYS

29. Broadside Array

An array is an aerial system built up from a number of elements, and arranged to give a field-strength diagram of some desired form. In the case of a broadside array the elements are arranged in a plane so as to give a sharply beamed lobe in a direction normal to the plane of the array. The elements are usually half-wave aeriels and such arrays are mainly used on metre wavelengths. As described in Sec. 4 a half-wave aerial radiates a spherical wavefront with the mid-point of the aerial as phase centre, i.e., as centre of the sphere. The field strength varies, at a given distance, according to $E_0 D(\theta)$, given in equation (8), in the plane of the aerial. In working out the field-strength diagram of an array it is convenient and correct to ignore this field strength variation in the first place and simply replace each half-wave aerial by an isotropic radiator or "point source" placed at the mid-point. The correction factor (8) can then be introduced when the field-strength diagram of the array of isotropic sources has been evaluated.

30. Linear Broadside Array

This is a broadside in which there is only a single row or column of elements fed in phase. The array may take four different forms

in practice according as the elements are horizontal or vertical and stacked vertically or horizontally. The four cases are illustrated in Figs. 777 to 780. Replacing the elements by point radiators as explained in the previous paragraph, one obtains either a row or column (i.e., a line), of point radiators.

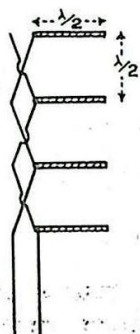


Fig. 777. - Vertical stack.

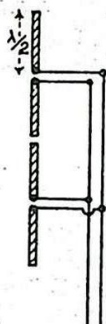


Fig. 778. - Vertical line.

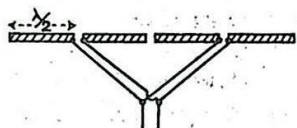


Fig. 779. - Horizontal line.

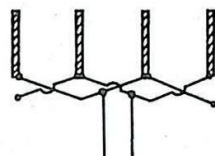


Fig. 780. - Horizontal row.

The field-strength diagram of such a line of isotropic radiators is obtained by adding up the waves, sent out in any direction, from all the radiators. In this section we shall limit discussion to arrays in which all the elements are equally energised.

Consider first the field-strength diagram in the plane through the line of radiators, i.e., the plane of the paper in Fig. 781. Take parallel rays going out from each radiator making an angle θ with the normal direction. Consider the ray from the centre radiator as standard if the number of elements is odd; if the number is even, introduce as standard a further hypothetical ray from the geometrical centre of the line of elements. Now the rays from the elements can be grouped in pairs on either side of the standard, one ray of the pair leading and the other lagging, by the same amount, on the standard. Such a pair is indicated by AB, CD in Fig. 781 with phase lead GF and phase lag CJ respectively. By the same argument as in Sec. 4 (See Fig. 748) the resultant field from such a pair has the same phase as the standard ray. Thus, the phase centre of the array is its mid-point. At a great distance it appears as if a spherical wavefront were spreading out from this point.

The variation of the strength of the wave as the direction θ varies is found by adding up all the rays. This is conveniently done simply by taking them in turn from left to right and adding vectorially. The vector addition can be done either graphically or algebraically. The result

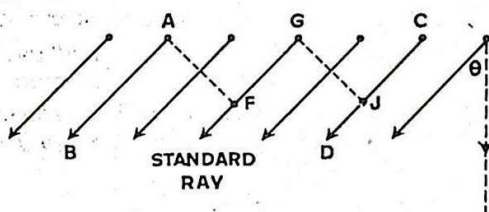


Fig. 781. - Line of isotropic radiators.

depends on the spacing between the point radiators. This spacing is nearly always $\lambda/2$ in practice, and the field-strength diagram is then given by :-

$$E \propto \frac{\sin [90^\circ n \sin \theta]}{\sin [90^\circ \sin \theta]} \dots\dots\dots (55)$$

where n is the number of elements in the array. Expression (55) is often called the Beaming Factor and sometimes the Grating Factor. Denoting it by $B(\theta)$ we have therefore :-

$$B(\theta) = \frac{\sin [90^\circ n \sin \theta]}{\sin [90^\circ \sin \theta]} \dots\dots\dots (56)$$

This is plotted for $n = 2, 4, 6$ and 10 in Figs 782, 783, 784, and 785. It must be multiplied by the field-strength factor of the individual element in this plane in order to obtain the actual field-strength distribution of the array.

The field-strength factor for the array, in a plane through the centre of the line of elements and perpendicular to its length (i.e. a plane through G in Fig. 781 and perpendicular to AGC is constant. There is no change in the field strength as one goes round the line of isotropic sources at a constant distance in this plane. In the case of the actual array, however, the field-strength factor of each individual element must be multiplied in.

Referring to the arrays illustrated in Figs. 777, 778, 779 and 780, it is now possible to obtain their field-strength diagrams in the two planes in terms of B , given by (56), and D , given by (9). The E-plane is taken as the θ -plane and H-plane as the ϕ -plane in accordance with Sec. 3.

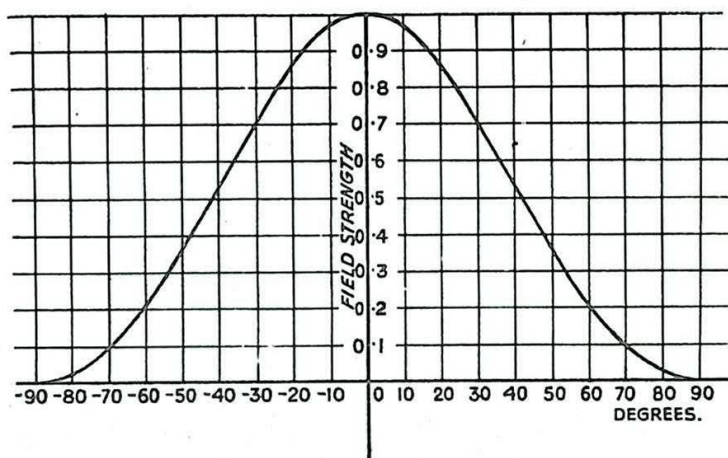


Fig. 782.- Beaming factor for two aeriels spaced $\lambda/2$ apart.

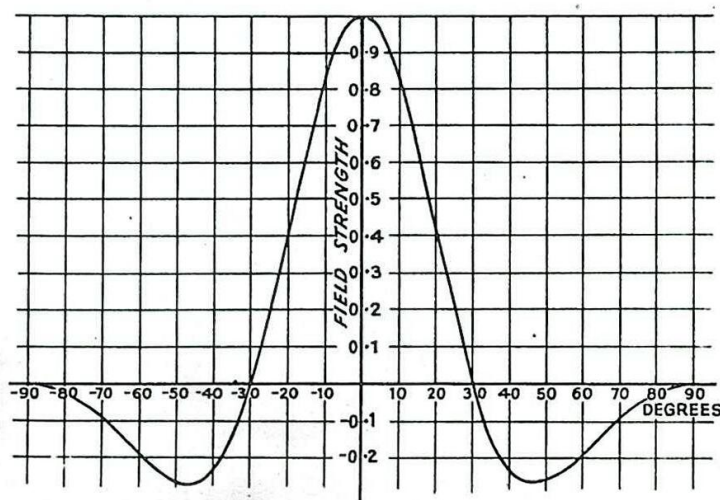


Fig. 783.- Beaming factor for four aeriels spaced $\lambda/2$ apart.

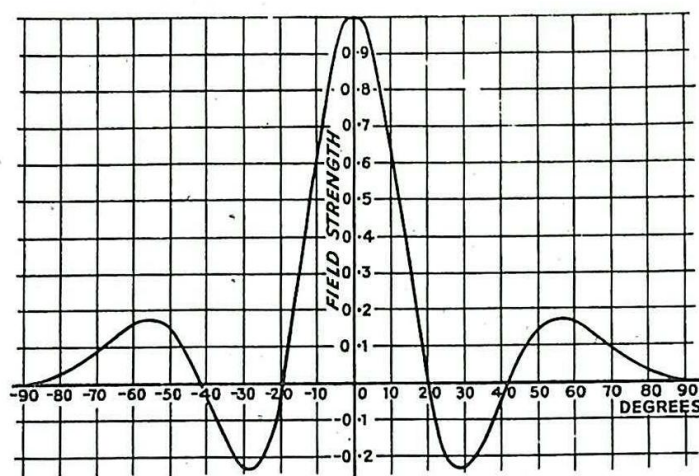


Fig. 784.- Beaming factor for six aerials spaced $\lambda/2$ apart.

<u>Figure</u>	<u>Description</u>	<u>Plane</u>	<u>Resultant Field-Strength Factor</u>
777	Horizontal Half-Wave Aerials in Vertical Stack with Half-Wave spacing.	Vertical or H-plane	$B(\phi)$ ($B(\phi)$ means that θ in equation (56) is replaced by ϕ .)
-do-		Horizontal or E-plane	$D(\theta)$
778	Vertical Half-Wave Aerials in Vertical Stack tip-to-tip.	Vertical or E-plane	$B(\theta).D(\theta)$
-do-		Horizontal or H-plane	constant
779	Horizontal Half-Wave Aerials in Horizontal Row tip-to-tip	Vertical or H-plane	constant
-do-		Horizontal or E-plane	$B(\theta).D(\theta)$
780	Vertical Half-Wave Aerials in Horizontal Row with Half-Wave spacing	Vertical or E-plane	$D(\theta)$
-do-		Horizontal or H-plane	$B(\phi)$

If elements other than half-wave aerials, but yet possessing axial symmetry, are employed, then the above table can be applied provided $D(\theta)$ is replaced by the appropriate field-strength factor. The main features of the beaming factors shown in Figs 782, 783, 784, and 785 are as follows :- There is a large maximum or Main Beam in the direction at right angles to the line of the array. The direction of this maximum

is sometimes called the Line of Shoot. Besides this there are a number of smaller maxima or Side Lobes.

The phase jumps 180° as the angle varies through the position of zero amplitude from one side-lobe to the next. An exception may occur between adjacent lobes at right angles to the main beam. The width of the main beam decreases as the number of elements increases. The Beam Width is usually defined as the number of degrees across the beam at half-maximum amplitude. This is plotted as a function of the number of elements in Fig.786. A rough rule, valid for a large number of elements, is

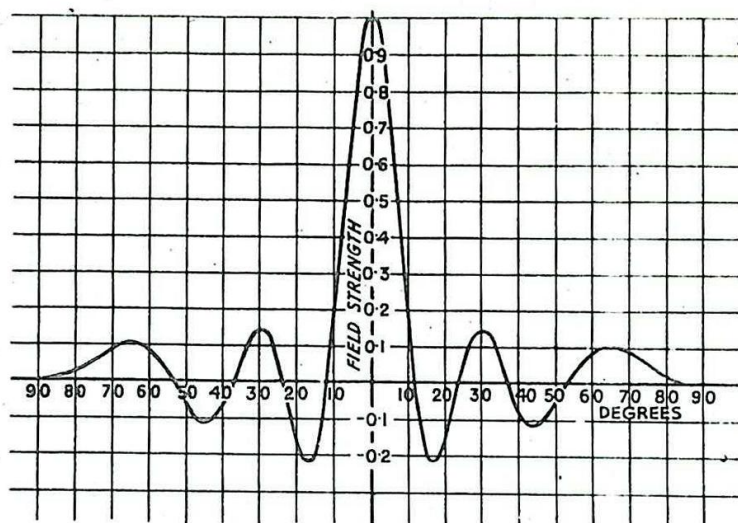


Fig.785.- Beaming factor for ten aeriels spaced $\lambda/2$ apart.

$$\text{beam width in degrees} = \frac{70 \times \text{wavelength}}{\text{width of array}} \dots\dots\dots (57)$$

The field-strength diagram of the individual elements, being rather broad, (Fig.749) hardly affects the beam-width when the number of elements is large.

When the number of elements is large, the first side-lobe is about 21% of the amplitude of the main maximum.

The gain (equation (65)), as can be shown by arguments given later, is about three times the length of the array expressed in wavelengths.

The choice of the aerial system for any particular function can be made from the table above. A

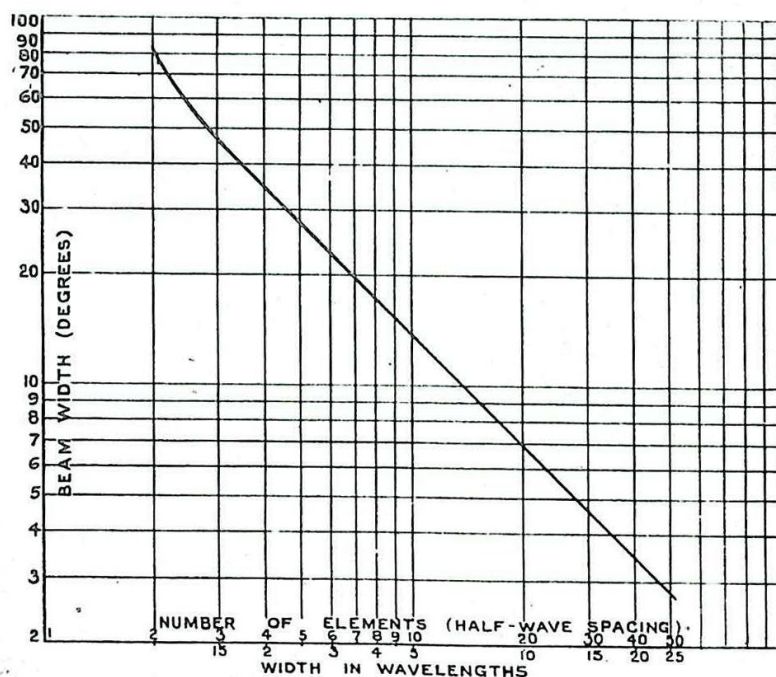


Fig.786.- Beam width of broadside.

beacon aerial, required to radiate all round and beamed in the vertical plane to prevent radiation being wasted by going vertically up into the sky would be best taken from Fig. 778. An aerial required to receive signals from a particular direction but from aircraft at any height would be as illustrated in Fig. 779 or Fig. 780

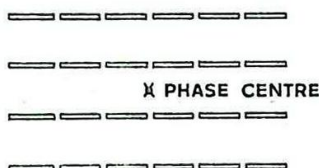


Fig. 787.- Complete broadside with horizontal polarisation.

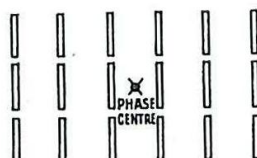


Fig. 788.- Complete broadside with vertical polarisation.

31. Complete Broadside Array

The complete broadside has elements in columns and rows fed equally and in phase. The elements will usually be half-wave aerials arranged either horizontally (Fig. 787) or vertically (Fig. 788), with centre spacing $\lambda/2$. To obtain the field-strength diagram of the broadside one first replaces the elements by point sources as in the previous paragraph. One then takes, say, a horizontal row of point sources and finds its beaming factor $B_h(\theta)$, by the method adopted for the linear array. Its phase centre is at its mid-point so that the whole horizontal row may itself be replaced by a point source at its mid-point radiating a spherical wavefront and with beaming factor $B_h(\theta)$ in the horizontal plane and constant in the vertical plane. Doing the same for all the horizontal rows, one finishes with a vertical line of point radiators. Assume now that these radiate isotropically. The phase centre is at the mid-point, i.e., at the centre of the whole array. The beaming factor in the horizontal plane is constant. In the vertical plane there is a beaming factor $B_v(\phi)$. For the whole array one takes the product of all the factors. Hence for horizontal polarisation we find :-

	Half-Wave Aerial	Horizontal Row	Vertical Row	Whole Array
Horizontal Plane	$D(\theta)$	$B_h(\theta)$	constant	$D(\theta) \cdot B_h(\theta)$
Vertical Plane	constant	constant	$B_v(\phi)$	$B_v(\phi)$

The corresponding results for vertical elements are obtained by reading "vertical" for "horizontal" and vice versa everywhere in the above table.

If there are n elements in a horizontal row and m elements in the vertical column, with $\lambda/2$ spacing between centres, then

$$B_h(\theta) = \frac{\sin(90^\circ n \sin \theta)}{\sin(90^\circ \sin \theta)} \dots\dots\dots (58)$$

and

$$B_v(\theta) = \frac{\sin(90^\circ \sin \theta)}{\sin(90^\circ \sin \theta)} \dots \dots \dots (59)$$

The field-strength diagrams of Figs 782, 783, 784 and 785 can thus be used to find the horizontal and vertical beaming factors of the complete broadside. When the beaming is great, with a large number of elements, the $D(\theta)$ factor of the individual elements hardly affects the beam width.

This type of aerial, the complete broadside, is used to obtain greater beaming in the two planes than is possible with a linear array.

32. Use of Reflecting Screen behind Broadside Array

The broadside arrays described above radiate both in front and behind as shown in Fig. 789. This is objectionable for most radar purposes and a reflecting metal screen may be positioned behind the array. This has the effect of blocking out the backward radiation. In order to investigate the effect of the screen in greater de-



Fig. 789.- Broadside without reflector (polar plot).

tail consider a single half-wave aerial a distance a in front of a metal sheet; (Fig. 790). As is well known, there can be no electric field along the surface of a good conductor (this is sometimes referred to as a boundary condition). Hence there can be no electric field along the line AB in Fig. 790.

Now remove the metal sheet and place a half-wave aerial a distance $2a$ from the first, fed with the same power but driven anti-phase (Fig. 791). At a point P on the line AB there is an electric field E due to each dipole and because of the anti-phase relationship these are as shown in the diagram.

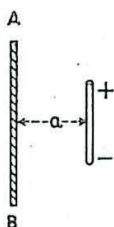


Fig. 790.- Aerial and metal reflector.

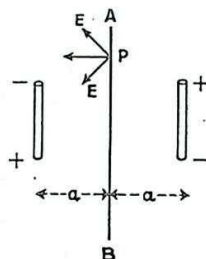


Fig. 791.- Aerial and image.

Their resultant is at right angles to AB, i.e. the electric

field due to this combination is zero along AB. Thus the field is zero along AB in both cases Fig. 790 and Fig. 791. By such arguments it follows that, so far as the field to the right of the screen is concerned, the two arrangements are equivalent. This is indeed just the principle of images, well known in optics and already discussed in Sec. 24.

Now consider the broadside array with a metal screen behind it. The phase centre of the broadside without the screen is at the centre of the array; (Sec 31). Replace the metal reflecting screen by an image source, anti-phase but of the same strength (Fig. 792). The radiated field to the right of the array with screen is obtained by multiplying the field-strength factor of these two anti-phase isotropic sources by the beaming factor of the array alone. The beaming factor of the array

is already known from Sec. 31. The field-strength factor of the pair of antiphase sources depends on their distance apart, $2a$. In practice a is usually $\lambda/8$ so that the source and its image are $\lambda/4$ apart. The field-strength diagram of such an arrangement can be given by the formula :-

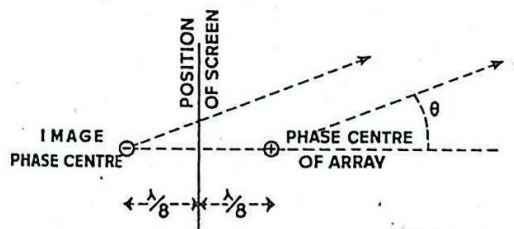
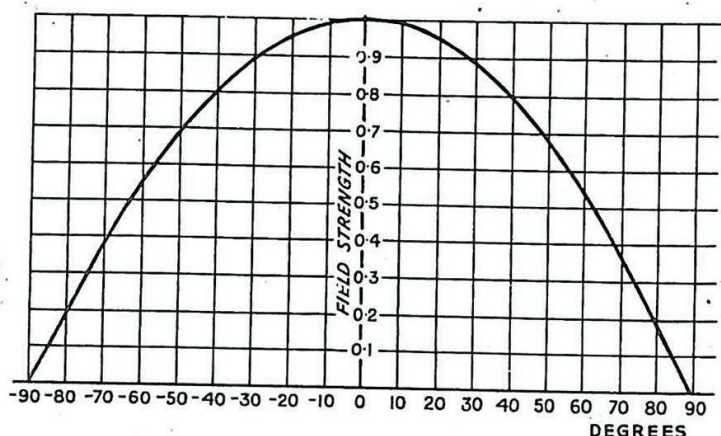


Fig.792. - Array with image.

Fig.793. - Beaming factor due to metal sheet placed $\lambda/8$ behind array.



$$E \propto \sqrt{2} \sin (45^\circ \cos \theta) \dots\dots\dots (60)$$

which is plotted in Fig.793. Due to this factor there is a slight sharpening of the beam, and a slight reduction in side lobes, because of the reflecting screen. The phase centre of the arrangement is in the plane of the screen. Formula (60) applies only in front of the screen, i.e., from $\theta = -90^\circ$ to $\theta = +90^\circ$. Behind the screen the field is always zero. Finally it is emphasised that the beaming given by (60) applies not only in the θ -plane but in all planes normal to the screen.

33. Wire Netting Reflector Screens

In practice, a complete metal sheet is objectionable since it would have a high windage area. The reflecting screen is normally made of wire netting. Some radiation thus leaks through the back of the screen. Suppose the screen to be made of wires placed vertically and horizontally, forming a square or rectangular network. It appears then that the wires effective in reflecting back the wave are those lying parallel to the electric field, and the spacing and thickness of these wires is the determining factor in judging how much radiation leaks through. Let the wires parallel to the electric field be of diameter d and spaced s apart; (Fig.794). Then the amplitude of the electric field leaking through the netting is given, approximately, as a fraction of the incident electric field, by the expression :-

$$\frac{2s}{\lambda} \cdot \log_{\epsilon} \left(\frac{s}{\pi d} \right) \dots\dots\dots (61)$$

the quantities s , λ and d all being expressed in the same units. The factor $2s/\lambda$ shows that the main requirement to ensure small leakage is that the spacing s of the wires shall be quite small compared with the wavelength. In addition, the logarithmic factor shows that thick wires are preferable to thin ones. Wire netting is often made with a hexagonal mesh and, due to the method of twisting together, some sets of wires are thicker than others; (Fig.795). Clearly this netting should be oriented so that the thick wires are parallel to the electric field.

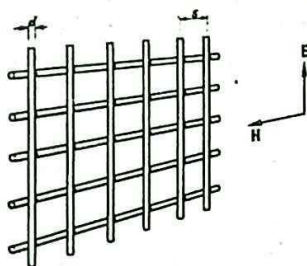


Fig.794.- Wave incident on wire netting.

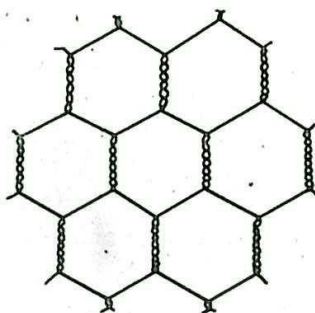


Fig.795.- Hexagonal wire netting.

As an example of formula (61), to give orders of magnitude, let us assume

$$s = \frac{1}{2}'' , d = \frac{1}{16}'' \text{ and } \lambda = 60'' . \text{ Then}$$

$$\frac{E(\text{leakage})}{E(\text{incident})} = \frac{2 \cdot \frac{1}{2}}{60} \cdot \log_{\epsilon} \left(\frac{\frac{1}{2}}{\pi \cdot \frac{1}{16}} \right)$$

$$= \frac{1}{60} \cdot (0.95)$$

$$= 1.6 \text{ per cent.}$$

If $d = 1/32''$, the answer is 2.7 per cent. The wire must be a reasonably good conductor for RF and thus iron wire should not be used unless coated over completely by some other metal such as zinc or tin.

In theory, a reflecting screen behind an array should be infinite in extent. If it is not sufficiently extended beyond the aerials at the sides of the array, radiation will leak round the edges of the screen and shoot out behind. In practice it appears that the screen need not extend more than $\lambda/2$ beyond the centre points of the outermost half-wave aerials, and indeed an extension of only $\lambda/4$ from the centres of these aerials is often found in practical installations, without, apparently, resulting in any appreciable loss.

34. Power Gain of Broadside Array

Since a broadside of half-wave aerials has a considerable beaming in its field-strength diagram, the power gain is high. The calculation of the power gain by the method of integration of power flow across a sphere is difficult. An approximate answer can be obtained as follows:-

Assume that R_r is the radiation resistance and I the RMS current, both measured at the centre of each half-wave aerial. (In practice all the aerials do not have the same radiation resistance due to mutual coupling. Neglect of this variation does not appreciably affect the accuracy in most cases). The power in each aerial is $I^2 R_r$ and, if there are N aerials altogether in the broadside, the total power P is given by

$$P = N I^2 R_r \dots\dots\dots (62)$$

At a distant point in the best direction, the fields from all the aerials in the array add up together. The array thus gives the same field at this point as would arise from a single aerial carrying current NI at its centre, i.e., supplied with power P' given by

$$P' = (NI)^2 R_r \dots\dots\dots (63)$$

Thus the power gain g_1 of the array relative to the half-wave aerial is given by

$$g_1 = P'/P \\ = N \dots\dots\dots (64)$$

The power gain g relative to an isotropic radiator will (by Sec. 10) be $\frac{1}{2}$ times this, i.e.,

$$g = 3N/2 \dots\dots\dots (65)$$

If a reflecting sheet is used behind the array, the gain is approximately doubled, so that finally, for a broadside of N aerials with reflecting screen,

$$g = 3N \dots\dots\dots (66)$$

The effective cross-sectional area A of an array of N aerials with $\lambda/2$ spacings is rather indefinite since the reflecting screen projects beyond the aerials. With the practical extensions of the screen discussed in Sec. 33, each aerial, including the outermost ones, is accounting for an area $\lambda/2$ square. Thus, assuming that this still holds for each aerial in the array, A is given by :-

$$A = N \left(\frac{\lambda}{2} \right)^2 \dots\dots\dots (67) \\ = N \frac{\lambda^2}{4}$$

Substituting for N in (66) and (67) we obtain

$$g = \frac{4 \cdot A \cdot 3}{\lambda^2}$$

A more accurate calculation, for a large broadside, gives

$$g = \frac{4\pi A}{\lambda^2} \dots\dots\dots (68)$$

and this latter is the formula which should be used. It will readily be apparent that these results depend rather much on the assumption of $\lambda/2$ centre-to-centre spacing between the aerials. It appears that this is a reasonable spacing to adopt and very small or very large spacings should be avoided. With close spacing, interaction between aerials becomes more pronounced and nullifies the increase in gain which would otherwise be expected. For spacing much greater than $\lambda/2$ side lobes are introduced with a consequent waste of energy.

The general idea of the "area approach" to broadside arrays should be noted :-

A given area or aperture A is available from space, mobility and other considerations. It should then be possible to obtain a power gain g given by the equation (68). This is done by filling the space with half-wave aerials placed about $\lambda/2$ apart and backing the space by a

reflecting screen; (Fig.796). The beam width in the horizontal plane is dependent on the horizontal width of the array, and in the vertical plane dependent on the vertical dimension.

35. Feeding Arrangements for Broadside Arrays

The first requirement in feeding a broadside array is to ensure that the elements are radiating in phase. This will be so if the elements are placed at intervals of $\lambda/2$ along the feeder, provided it is an open wire feeder and that the feeder is crossed over between each aerial. The velocity of propagation in polythene-filled feeders is about $2/3$ that in free-space and λ in such a feeder is $2/3$ of the free-space wavelength, and allowance must be made for this. The next requirement is to obtain reasonable matching between the aerials and feeders. Centre-fed half-wave aerials are unsuitable because, when several are paralleled, the load on the feeder is small ($73/N$ ohms, where N is the number of aerials.) Most open wire feeders have a characteristic impedance of from 300 to 600 ohms. It is therefore more convenient to end-feed the half-wave aerials (Sec. 11) and so obtain loads of the order of the feeder impedance. A slight mismatch can be removed by the use of stubs. In practice, therefore, broadside arrays are often built up in bays - each bay consisting of two stacks of half-wave aerials, end-fed (Fig.797). The bays must all be fed in phase and with the same amount of power. This is ensured by careful cutting of the feeder lengths and by matching each bay to its feeder.

The input resistance of a pair of half-wave aerials has been given in Sec. 11. However, due to coupling between the aerials and the proximity of the reflecting screen, the input resistance may differ considerably from the free-space value. This neglect of interactions between the aerials affects the theoretical arguments of Sec. 34, but the final results (65) and (68) are probably not far wrong owing to compensating factors.

36. Example of a Complete Broadside Array

A broadside array measures $25' \times 10'$ and consists of 40 half-wave aerials arranged with $\lambda/2$ spacing between their centres, the wavelength being $5'$. The array is backed by a metal reflecting screen. Find the power gain and the beam widths in the two principal planes.

From (68) we have

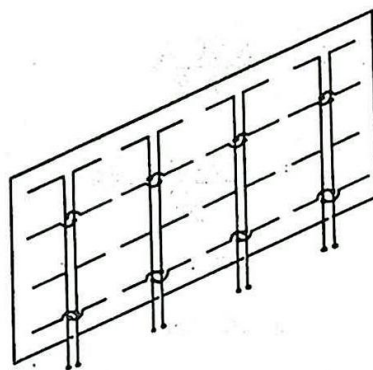


Fig.796.- Schematic diagram of broadside array backed by reflector sheet.

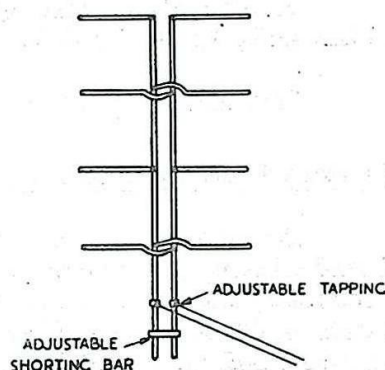


Fig.797.- Bay of half-wave aerials.

$$\begin{aligned}
 g &= \frac{4 \pi A}{\lambda^2} \\
 &= \frac{4 \cdot \pi \cdot 25 \cdot 10}{5^2} \\
 &= 126 \\
 &\text{or } 20 \text{ db.}
 \end{aligned}$$

In the plane through the long dimension there are ten aerials in a line and hence from (57)

$$\begin{aligned}
 \text{beam width} &= \frac{70.5^\circ}{25} \\
 &= 14^\circ.
 \end{aligned}$$

Alternatively the same result may be obtained from Fig.786.

In the plane through the short dimension there are four aerials in a line and, from Fig.796, the beam width = 34° .

$$\text{Formula (57) gives } \frac{70 \times 5}{10} = 35^\circ.$$

37. Tapered Feed to Linear Broadside Arrays

Returning to a consideration of a line of radiators (see Sec.30) the field-strength diagrams shown in Figs.782, to 785 have an objectionable feature, namely the small side lobes on either side of the main beam. Sometimes difficulties and mistakes arise in radar due to signals being transmitted and received back via the side lobes, as well as by the main lobe. Side lobes may be eliminated by Tapered Feeding which may be explained as follows :-

Take two half-wave aerials placed $\lambda/2$ apart and fed equally and in phase. Neglecting the field-strength factor of the individual aerials, which may be introduced as a correction factor at a later stage, the field-strength diagram is as shown in Fig.782. There are no side lobes, but the main beam is, of course, very wide. The field-strength factor obtained from (56) with $n = 2$, is

$$\frac{\sin \left(\frac{180^\circ \sin \theta}{90^\circ \sin \theta} \right)}{\sin \left(\frac{90^\circ \sin \theta}{90^\circ \sin \theta} \right)} \dots\dots\dots (69)$$

A plot in polar co-ordinates is shown in Fig. 798. The phase centre X is half-way between the aerials, and they may be replaced by a source positioned at that point. Now take another similar pair of aerials fed with the same currents as the first, and place the two pairs together as shown in Fig.799. Each pair is replaceable by a source at its mid-point so that the arrangement of Fig.799 is equivalent to that of Fig.780, i.e. two sources $\lambda/2$ apart. The field-strength diagram of these two sources is the same as the original. Hence the resultant field-strength diagram of the arrangement of Fig.799 is equivalent to that of three aerials fed with currents 1, 2, 1 units respectively (Fig.801).



Fig.798.- Field-strength diagram of two aerials spaced $\lambda/2$ apart and fed in phase.

The phase centre of these three aerials is at their mid-point. The resultant beaming factor is the square of that given in Fig.782, for the two aerials only. Continuing this process, combine 1, 2, 1 units with another 1, 2, 1 units placed so that the mid-points or phase centres are $\lambda/2$ apart. Then the original beaming factor (69) again arises from their pair and the resultant beaming factor is the cube of the original one. This arrangement is equivalent to four aerials fed with currents 1, 3, 3, 1 units. The phase centre is at the mid-point.

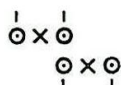


Fig. 799.- Combination of two pairs.



Fig.800.- Replacement of pairs by sources at phase centres.



Fig.801.- Two pairs combined to form three aerials fed in ratio 1:2:1.

It is clear that the process can be extended and the required currents in the aerials will be as shown:-

<u>Number of Aerials</u>	<u>Currents</u>	<u>Beaming Factor</u>
2	1, 1	$T(\theta) = \frac{\sin(180^\circ \sin \theta)}{\sin(90^\circ \sin \theta)}$
3	1, 2, 1	$[T(\theta)]^2$
4	1, 3, 3, 1	$[T(\theta)]^3$
5	1, 4, 6, 4, 1	$[T(\theta)]^4$
6	1, 5, 10, 10, 5, 1	$[T(\theta)]^5$
$n + 1$	$1, n, \frac{n(n-1)}{1.2}, \frac{n(n-1)(n-2)}{1.2.3}, \dots, n, 1$	$[T(\theta)]^n$

As the original beaming factor is raised to higher powers, the beam becomes sharper. For example, consider eleven aerials fed with currents

1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1,

and spaced $\lambda/2$ apart. The beaming factor of this arrangement is

$$\left[\frac{\sin(180^\circ \sin \theta)}{\sin(90^\circ \sin \theta)} \right]^{10}$$

and the corresponding field-strength diagram is plotted in Fig.802. There are no side lobes but the beam is much wider than one would obtain with a similar number of aerials fed with equal amplitudes. Comparison with Figs.783, 784 and 785 illustrates the effect. The beam width at half

amplitude in Fig.802 is 26° , whereas if eleven equally fed aerials were used the beam width would be (from formula (57) or Fig.786) about 62° . If the two end aerials of the present arrangement were omitted the field-strength diagram would not be altered much, since they are carrying so little current, but even then (nine elements) the beam width is much greater than that for the same number of equally fed aerials.

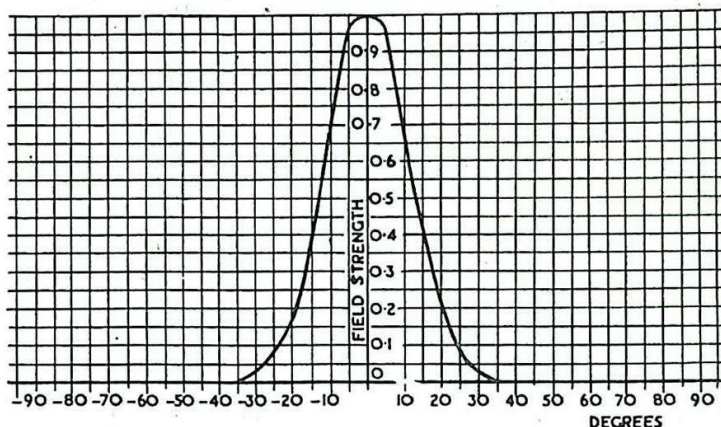


Fig.802. - Beaming factor for eleven aerials spaced $\lambda/2$ and with binomial distribution.

The method of feeding described above involved high

currents in the centre elements, tapering off to small values at the end elements. It is therefore called Tapered Feed. Since the numbers representing the currents in the elements are also the coefficients of x , in the expansion of $(1 + x)^n$, the distribution of currents is sometimes called a Binomial Distribution. (As n becomes increasingly large the binomial tapering tends towards the Gaussian distribution familiar in the theory of probability).

At the moment of writing this type of array is not in use, but it may be found in future developments at very short wavelengths. In any case, the principle is of importance in connection with the illumination of mirrors (see Sec. 48).

38. Beam Swinging with Linear Broadside Array

Instead of feeding the elements in phase as in Sec. 30 one may introduce a progressive phase difference in the radiating current elements. This may be done to any desired amount by having adjustable line-lengtheners in the feeder leads. Consider a set of radiators in a line, spaced

$\lambda/2$ apart as in Sec. 30.

Suppose that each radiating element has the same amplitude but that the phase

of each element lags 90° on the element on its immediate left-hand (Fig.803) Take rays going out from each element at an angle of 30° with the normal to the line of radiators. The ray from A in Fig.803 has a spatial lag of amount AC on the ray from B. Since $AB = \lambda/2$ and CAB is a 30° right-angled triangle with sides in the ratio $1:2:\sqrt{3}$, the spatial lag AC is $\lambda/4$, i.e. 90° of phase lag. However, A is being driven 90° of phase in advance of B, so that on the whole these rays from A and B arrive in phase at a distant point. The main beam of the array is thus directed at an angle of 30° to the normal, due to this method of feeding with a progressive phase advance of 90° . In the case of a progressive phase difference

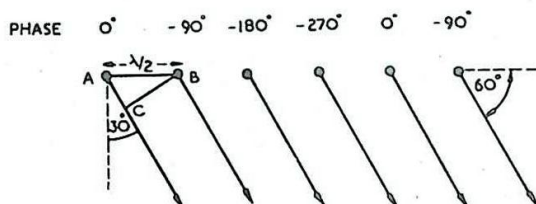


Fig.803. - Swinging of beam by progressive phase lag.

of ϕ degrees between neighbouring elements the angle α which the beam makes with the normal, is given by :-

$$\sin \alpha = \phi/180 \dots\dots\dots (70)$$

This applies when the centres of the elements are spaced $\lambda/2$ apart. When the spacing is not $\lambda/2$, but, say, λ/x , then α is given by

$$\sin \alpha = \frac{x\phi}{360} \dots\dots\dots (71)$$

The complete field-strength diagram is given in Fig. 804 for the case of six aerials.

39. End-Fire Arrays

If the phasing of the elements of a linear array is arranged so as to swing the beam round through 90° from the normal, an End-Fire Array is obtained. For this to occur, with $\lambda/2$ spacing, the phase difference between the currents in successive elements must be 180° . For spacing λ/x , the phase difference between driven elements must be equal to the phase

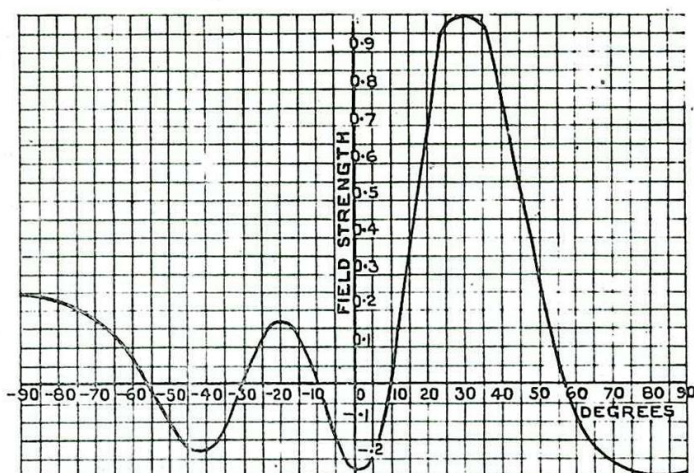


Fig. 804. - Beaming factor for six aerials spaced $\lambda/2$ apart and fed with a progressive phase change of 90° .

corresponding to the spacing, so that the rays going out from the elements in a direction along the line of the array are in phase; (Fig. 805)

The optimum spacing of the elements is probably something less than $\frac{3\lambda}{8}$; and the corresponding gain relative to a single element is

about three times the length of the array expressed in wavelengths. If the elements are half-wave aerials, each of gain $1\frac{1}{2}$, the overall gain g relative to an isotropic radiator is

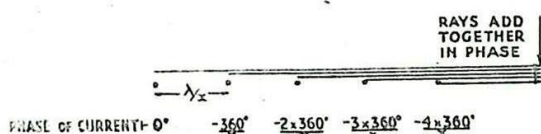


Fig. 805. - End-fire array.

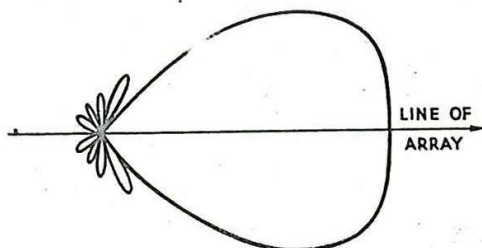


Fig. 806. - Beaming of end-fire array with ten elements spaced $\lambda/4$ apart.

$$g = 4.5 \text{ (Length of End-Fire Array in Wavelengths) } \dots\dots (72)$$

A typical field-strength diagram is shown in Fig.806. The side lobe is about 21% of the main beam. A rough rule for the beam width at half amplitude is

$$\frac{200 \cdot \text{wavelength}}{\text{length of array}} \text{ degrees } \dots\dots\dots (73)$$

The beaming takes place in both E- and H-planes, in contra-distinction to the linear broadside array.

40. Comparison of End-Fire and Broadside Arrays

The most notable feature of the end-fire array is the blunt end of the field-strength diagram (Fig.806) compared with that for a broadside. Consider a linear array with elements fed in phase. Rays normal to the array are in phase. As soon as one takes a direction slightly inclined to the normal, appreciable phase differences arise (Fig. 807), and the resultant field strength is diminished from the value in the normal direction. Now consider an end-fire array. The rays are in phase along the array. If one considers rays inclined at a slight angle to the line of the array (Fig.808) hardly any change in the phase difference is

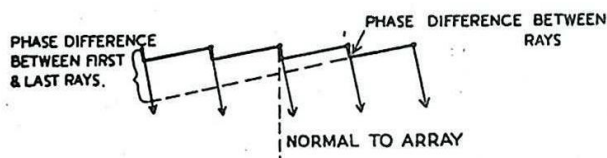


Fig.807.- Phase difference introduced by slight deviation from normal.

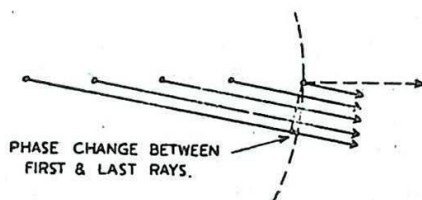


Fig.808.- Phase change introduced by slight deviation from line of array (phase change between successive rays is too small to show in this diagram).

introduced. The deviation from the line of the array must be quite large before these phase differences become appreciable. Hence the end-fire array produces a field-strength diagram with a blunt end. The lobe can be sharpened by making the phase of the currents in successive elements differ by an amount greater than the physical spacing of the elements. Then in the direction of the line of the array rays are already out of phase, and for the same deviation as before a greater phase change is obtained. The main beam is thus sharper, but since the maximum is not now formed by adding rays in phase, the ratio of side to main lobes is increased compared with what one would obtain from an end-fire array fed in the usual way.

41. Parasitic Director

One of the troubles of an end-fire array lies in arranging the feed of the elements in correct phase. This can be overcome to some extent by using parasitic radiators in line with a single active aerial. Consider, for example, one $\lambda/2$ aerial connected to a transmitter and a second

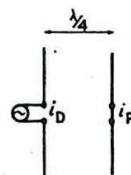


Fig.809.- Driver and parasite.

aerial, about $\lambda/2$ long and short-circuited at its centre, placed $\lambda/4$ away, as in Fig. 809. This second aerial is excited by the fields associated with the first, and is called a Parasite, the other being termed the Driver. Suppose the current in the driver is i_D . The field radiated from the driver is lagging 90° on i_D (see Sec. 4). In travelling out from the driver to the parasite the electric field suffers a further phase lag of 90° due to covering the distance $\lambda/4$. Hence the electric field at the parasite, or the EMF v induced in the parasite, is 180° out of phase with the driver current i_D . The parasite is regarded as a receiving aerial picking up from the driver. Owing to the fact that no load is attached to the parasite (its terminals being short circuited), all the power picked up is re-radiated. If the current at the centre of the parasite is i_P and the radiation impedance z_P , then using (29) in Sec. 14 with $z_\ell = 0$, we find

$$i_P = \frac{v}{z_P}$$

If the length of the parasite is about 0.48λ , its radiation resistance will be about 73 ohms and there will be no reactance in z_P . If, however, its

length is less than 0.48λ there will be a capacitive term in z_P ; (Fig. 810).

Thus by shortening the parasite, i_P may be made to lead on v . If the

length is suitably chosen a phase lead of 45° may be obtained. The current

in the parasite then lags on the current in the driver by 135° as shown in Fig. 810. This is seen to be an end-fire type of arrangement as

discussed in Sec. 40. The phase of the current lags by more than the amount corresponding to the physical spacing of the elements. The field-

strength diagram may be derived from first principles by considering the addition of wavelets from the driver and parasite; (Fig. 811). At a

distant point along the line AB of the elements, the driver field leads the parasite field by 45° due to the driver's spatial lag of 90° ($\lambda/4$) and current lead of 135° . However,

as one moves round from the line AB the spatial lag becomes less, and thus the field at a distant point from the driver leads more and more on the field from the parasite, so that the resultant field gets less and less. At an angle of 120° from AB, the driver and parasite fields are anti-phase and the field is a minimum. The field-strength

diagram in the H-plane is of the form shown in Fig. 812 (polar plot) with a main lobe in the line of the

elements and a small back lobe. There is a minimum at 120° but it is not usually a zero in practice since the current in the parasite is less than the current in the driver and the two fields do not cancel.

A parasite used in this way, cut shorter than resonant length, is called a director, since it helps to direct the radiation in the direction from driver to parasite.

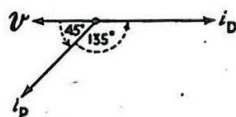


Fig. 810.- Phases of currents in driver and parasite and E.M.F. in parasite when acting as a director.

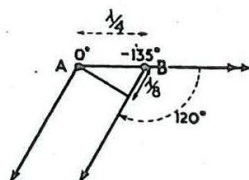


Fig. 811.- Parasite acting as a director.

The length of the parasite required to provide a 45° phase difference between EMF and current in the parasite may be obtained from the theory of resonance applied to the aerial as a tuned circuit.

From the results given in Chap. 1 Sec. 19 ,

$$\tan \phi = 2 Q \delta$$

where ϕ is the phase difference and δ the fractional detuning; it follows that for ϕ to be 45° , $2Q\delta$ must be unity. For a typical "half-wave" aerial of resonant length 0.48λ the ratio of diameter to wavelength is of the order of $1/1000$, so the Q of the resonant circuit is approximately 7 (see Sec. 12). The fractional detuning δ must therefore be equal to $1/4$ so that the change in length is

$$\frac{0.48\lambda}{14} = 0.035\lambda.$$

As mentioned in Sec. 12 the length must be shorter than that required for resonance if the impedance is to be capacitive. Hence the required length is

$$\begin{aligned} 0.48\lambda - 0.035\lambda \\ = 0.445\lambda \end{aligned}$$

The current i_p then leads the EMF v by 45° .

Although a spacing of $\lambda/4$ was used in the argument above, other spacings may be employed. Depending on the length of parasite and the spacing, a variety of field-strength diagrams may be obtained. Normally, it is as well to have the parasite fairly near to the driver in order to obtain a large current in it.

42. End-Fire Array with Parasitic Directors (Yagi)

By using a driver and several parasites of the type described in Sec. 41 an end-fire array may be constructed; (Fig.813). Generally speaking; the phase lag of the current in any element relative to that in its left hand neighbour is greater than the phase difference corresponding to the element spacing in wavelengths. Hence this is not, strictly speaking, a simple end-fire array but is of the type described in Sec.40.

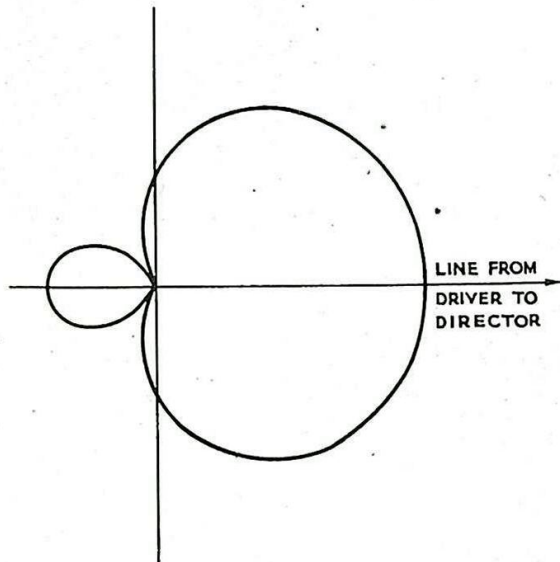


Fig.812. - Field-strength diagram of driver and director.

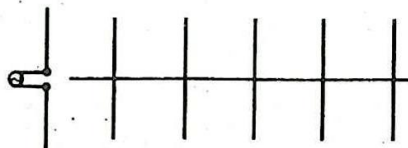


Fig.813. - Yagi aerial.

The side lobes are large and the main beam fairly sharp.

This type of array is called a Yagi Aerial. The gain g of a Yagi is probably not much different from that of a simple end-fire array, the sharpening of the beam being compensated for by the unwanted power in the side lobes. Thus using equation (62), we have, approximately,

$$g = 4\frac{1}{2} \text{ (Length of Yagi in Wavelengths) } \dots\dots\dots (74)$$

Higher gains than this are occasionally obtained by carefully varying the spacing and length of the directors, particularly with short Yagi aerials. When the Yagi is adjusted for maximum gain, the side lobes are about 30% of the amplitude of the main beam and the beam width at half maximum amplitude is roughly given by

$$\text{beam width in degrees} = \frac{130}{\text{Length of Yagi in Wavelengths}} \dots\dots (75)$$

It is emphasised that the Yagi beams in both E- and H-planes and is comparable in this effect with either a complete broadside or a linear end-fire array.

The spacing between directors is not critical and for a given spacing the director length is adjusted by trial and error to give the best result. The parasites may all be supported at their centres by an earthed metal rod (Fig. 813) since the voltage at the centre of a parasite is zero.

43. Parasitic Reflector

In Sec. 41 the effect of cutting the parasite less than resonant length was discussed. It led to the result that the parasite current was lagging on the driver current. By a similar argument it can be shown that the current in a parasite in the same position as before but slightly longer than resonant length is leading on the driver current, and the maximum of the field-strength diagram is in the direction from parasite to driver. The long parasite thus appears to reflect the wave back in the direction of the driver and is called a parasitic reflector. It is also called a tuned reflector since it must be cut to a special length in order to work properly.

Parasitic reflectors are sometimes used in place of metal or netting screens behind broadside arrays; each element of the broadside has a parasitic or tuned reflector behind it. In contra-distinction to the tuned reflector a netting screen is called an aperiodic reflector.

Very often a Yagi aerial has, in addition to several directors, a reflector behind the driver. This reflector may be either a tuned parasitic reflector or a netting screen.

44. Use of Folded Half-Wave Aerial

Due to the proximity of the parasites the input impedance at the centre of the driver of a Yagi is usually low, sometimes as small as 20 ohms. Transmission lines seldom have a characteristic impedance much less than 80 ohms so that a quarter-wave matching transformer or other matching device is required between driver and feeder. The use of a folded dipole enables the transformer to be dispensed with.

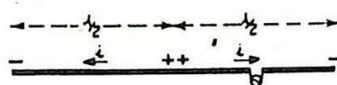
Take a centre-fed half-wave aerial with current i at its feed point. Extend the left arm a further half wavelength as shown in Fig. 814. By analogy with transmission lines there is a current $-i$ at the centre of the added piece and the voltages are as shown. Bend the added piece round so as to be very near the original aerial. We now have two components, each i

Since the folded part is very near the driving part, the power radiated is practically the same as that for a half-wave aerial carrying current $2i$. Let R_r be the radiation resistance measured at the input to the folded aerial. The radiation resistance of a half-wave aerial is 73 ohms. Thus, equating powers, we have :-

$$(2i)^2 73 = i^2 R_r$$

$$\therefore R_r = 4.73$$

$$= 292 \text{ ohms}$$



(a) BEFORE FOLDING



(b) AFTER FOLDING

Fig.814.- Folded dipole.

In general the input impedance is quadrupled and thus a low input impedance is brought up to a suitable value for direct matching to a transmission line.

45. Bandwidth of Arrays

In Sec. 40 a discussion was given of the Q-Factor of a single half-wave aerial. The calculation of the Q-factor or band-width of a large array is an extremely difficult task. The aeriels in an array are inter-coupled both by their fields and by the connecting feeder runs. Generally speaking, therefore, an array can seldom be used on any except a narrow frequency band. The following example gives an idea of the orders of magnitude; A six-element array, in use, working on 30 Mc/s. can operate at frequencies within about 500 kc/s. on either side of the design frequency. Beyond this the operational efficiency is very noticeably reduced. At high powers the maximum deviation is as low as 200 kc/s. owing to sparking and brushing which arise on the feeders when the frequency is changed. The elements of this array are constructed of thin wire ($\lambda/\text{diameter} \approx 1500$). To improve the performance thicker tubing might be used, or a really thick aerial could be simulated by a wire cage type of construction. The sparking difficulty could be eliminated only by widening the feeder spacing.

46. Effect of Ground on Aerial Arrays

In Sec. 24 it was shown that a single aerial placed h above ground could be replaced by two aeriels separated by $2h$ energised anti-phase in free space, and the resultant interference pattern obtained.

Consider now an array of horizontal aeriels seen end-on in Fig. 815. Each aerial A_1, A_2, \dots can be associated with its image A_1', A_2', \dots and the image of the array is thus formed, as shown, to take account of the presence of the earth.

In the general case illustrated in Fig. 815 the free-space field-strength diagram of the aerial system is not isotropic and the method of replacing each system and its image by an isotropic radiator as explained previously cannot be used. Usually, however, any aerial systems in which ground reflections are important are set up so that the field-strength diagram of the array (neglecting the effect of ground) is symmetrical above and below the horizontal (e.g. a broadside array when used

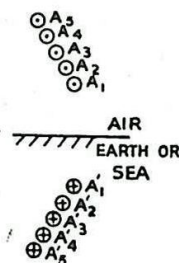


Fig.815.- Image of array with horizontal polarisation.

in ground radar systems). In this case the field strength is the same for both direct and reflected rays at all angles of incidence and for the purpose of finding the effect of the ground the system and its image may be replaced by isotropic sources at their phase centres. The free space field-strength diagram of the system can be introduced later as a correction factor. The table in Sec. 26 applies to any array under this condition of symmetry about a horizontal plane.

47. Gap-Filling by Means of Phasing

Suppose an aerial system is split into upper and lower halves which are normally fed in phase. In order to fill the gaps in the vertical field-strength diagram, a switching arrangement is introduced in the feeders whereby the halves may be fed antiphase. Consider the aerial system AD, Fig. 816, made of two identical sub-systems AB and CD. The field-strength diagram in free space of either of the sub-systems is a graph which might, for example, be one of the graphs of Figs. 782, 783, 784 or 789. Let P_1

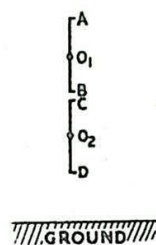


Fig. 816.- Aerial system split in halves.

be the corresponding beaming factor. Suppose the sub-systems to be fed in phase and imagine them replaced by isotropic radiators at the phase centres O_1, O_2 . For the sake of argument suppose the distance between O_1 and O_2 is 2λ . The field-strength diagram of the pair of isotropic radiators is shown in polar co-ordinates in Fig. 817. Let P_2 be the corresponding beaming factor. The beaming factor of the aerial system AD is the product $P_1 P_2$. There is a main beam in a direction parallel to the earth's surface and the whole diagram is symmetrical above and below the horizontal.

Reflections take place from the earth and interference lobes arise in the way explained in sections 24 to 26 and 46.

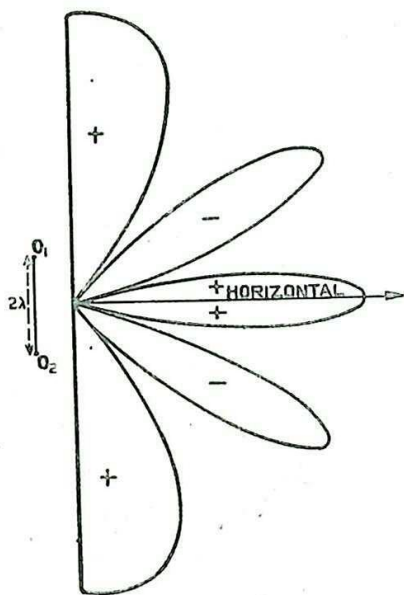


Fig. 817.- Field-strength diagram of O_1 and O_2 when fed in phase.

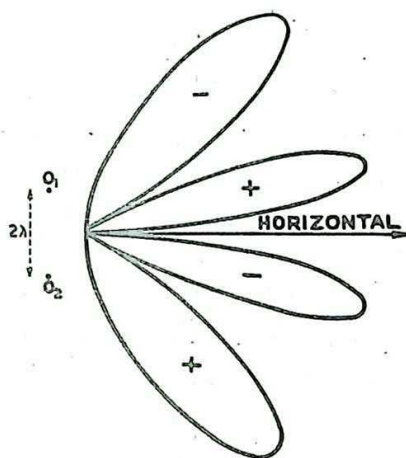


Fig. 818.- Field-strength diagram of O_1 and O_2 when fed in anti phase.

Next consider the sub-systems AB, CD to be fed antiphase. Replace each sub-system as before by an isotropic radiator, the two radiators this time being fed anti-phase. The resultant beaming factor is P'_2 , giving the field-strength diagram shown in Fig. 818. The

beaming factor for the complete system AD ignoring the effect of ground is the product $P_1 P_2'$. Whatever the nature of P_1 there will be no radiation in the horizontal direction since P_2' in this direction is zero. In fact the zeros of P_2' correspond to the maxima of P_2 and vice versa; also the corresponding lobes above and below the horizontal are of opposite phase for P_2' and in phase for P_2 . Hence where these interfere after reflection from the ground to cause zeros in one case they are additive in the other.

Thus irrespective of the nature of P_1 the change in the method of feeding the two halves of the system from in-phase to anti-phase causes the gaps due to ground reflection to appear where lobes appeared before and vice versa. At the same time the main line of shoot is tilted up at an angle x_0 depending on the separation of the centres of the upper and lower halves of the aerial. For a separation λ , $x_0 \approx 30^\circ$; for 2λ $x_0 \approx 14\frac{1}{2}^\circ$; for 3λ , $x_0 \approx 9\frac{1}{2}^\circ$; 4λ , $x_0 \approx 7\frac{1}{4}^\circ$ and so on.

MICROWAVE AERIALS

48. Paraboloidal Mirrors

It has been emphasised in Sec. 34 that in order to attain a narrow beam and high gain, the radiating sources must be spread over as large an area A as possible. In microwave work for wavelengths of 50 cms or less, the paraboloidal mirror (Fig.819) enables the effect of distributing the sources to be attained without the use of more than one radiator. The mirror is made of metal such as aluminium alloy or copper; sometimes wire netting is employed. A section of the mirror is shown in Fig.820. The action of the mirror is similar to that which occurs in ordinary optics in the case of reflecting paraboloids.

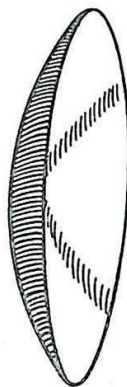


Fig.819.- Paraboloidal mirror.

The important point relating to a parabola is its focus F. If rays of electromagnetic radiation strike the mirror, all being parallel to the axis FO, the rays converge at F. A focal plane MFN through F and perpendicular to OF usually coincides with the aperture of the mirror. Suppose now that a radiating element is placed with its phase centre at F. Draw rays going out from F and being reflected at the surface of the mirror. After reflection the rays are all parallel to the axis OF. Further, if we consider the focal plane MFN, then all rays starting out from F reach this plane in the same time; i.e., in Fig.820 all the paths FAB, FHG, FKL are the same length. Thus, F being the phase centre of the radiator, the rays crossing the plane MFN are all in phase so that MFN is an equiphase surface. We can regard this equiphase surface as constituting a radiating area A filled with radiating sources, and the field-strength diagram, etc., can be determined from a knowledge of these sources just as in the case of a broadside array.

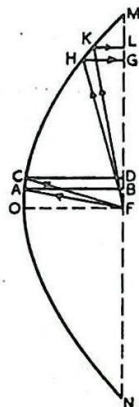


Fig. 820.- Section of paraboloid

The strengths of these effective sources distributed across the circular aperture MFN depend on the field-strength

diagram of the primary source at F. If the primary source produces a uniform distribution of field strength across the aperture then the results are comparable with what one obtains in the theory of broadside arrays. There is a main lobe with side lobes and the gain g is given by formula (58)

$$g = \frac{4\pi A}{\lambda^2} \dots\dots\dots (76)$$

The beam width at half amplitude is approximately

$$80 \lambda/d \text{ degrees} \dots\dots\dots (77)$$

where d is the diameter of the mirror in the aperture plane (cf. formula 57). The first side lobes on either side of the main beam are 13% of the maximum amplitude (compare Secs. 30 and 31 which give 21% side lobe).

Thus with the circular aperture, the beam is wider and the side lobes less than from an equal rectangular aperture, but the gain is the same.

A uniform distribution across the aperture is never obtained in practice due to the peculiar field-strength distribution which would be required from the primary source at the focus. For example, referring to Fig. 820, rays FA, FC have been drawn enclosing an angle of 5° and resulting in the illumination of a region BD in the aperture plane. Another pair of rays FH and FK enclosing an angle of 5° illuminate a region GL in the aperture plane. It is clear from the figure that LG is about $1\frac{1}{2}$ times BD. Thus, if the aperture is to be equally illuminated, less power must be sent out from the primary source at F in the direction FA (or FC) than in the direction FH (or FK). Indeed the primary source must have a distribution such that the field strength (at a given distance) in the direction towards the edge of the mirror is twice that (at the same distance) in the direction of the centre of the mirror. The power density at the edge must be four times that at the centre.

Most primary sources give less field strength towards the edge of the aperture plane than at the centre. The feed is thus tending to be tapered (cf. Sec. 37) and this results in a reduction of side lobes and broadening of the main beam in the field-strength diagram from the mirror. For a paraboloidal mirror (circular aperture) of diameter d the following data have been worked out assuming the distribution of the field strength over the aperture to be part of a Gaussian curve (see Sec. 37):

Amplitude of field at edge of Aperture divided by Amplitude at centre of Aperture	Beam Width right across between first zeros (degrees)	Beam Width right across at half max. amplitude (degrees)	Power Gain g	Amplitude of First Side Lobe expressed as percent- age of main beam
1.00 (uniform distribution)	140 λ/d	80 λ/d	$1.00(\frac{\pi d}{\lambda})^2$	13%
0.37	162 λ/d	89 λ/d	$0.92(\frac{\pi d}{\lambda})^2$	8.5%
0.13	202 λ/d	100 λ/d	$0.76(\frac{\pi d}{\lambda})^2$	2.5%

These figures show well the broadening of beam, diminution in gain and reduction in side lobes associated with concentration of the distribution towards the centre of the aperture plane.

49. Primary Feeds for Paraboloidal Mirrors

(i) Rear Feed with Half-Wave Aerials

This is illustrated in Fig. 821. A coaxial feeder passes through the centre of the mirror and ends in a half-wave aerial with parasitic reflector. The field-strength diagram of the primary source is different in the E- and H-planes. In the E-plane more energy is concentrated towards the centre of the mirror and the beam from the mirror is broader in this than in the H-plane. Unless a balance-to-unbalance transformer is used RF currents flow on the outside of the coaxial

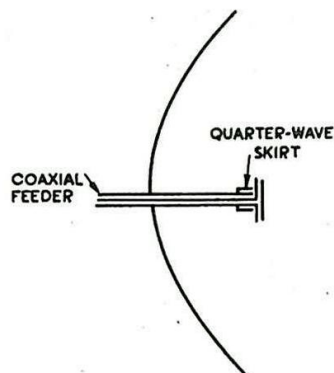
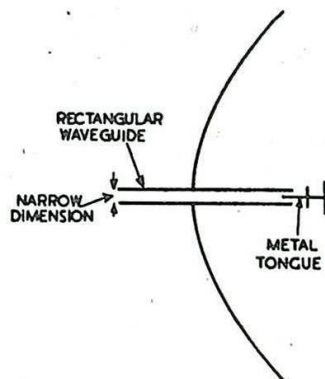


Fig.821.- Half-wave aerial with reflector for feeding parabolic mirror.

due to one leg of the aerial being attached to this outer conductor. A quarter-wave can (skirt or bazooka) or other device is therefore attached which inserts an infinite impedance between the end of the coaxial outer conductor and the end of the can. This decouples the rest of the outer conductor from the feed point to the aerial (see Chap. 4 Sec. 38). Instead of a parasitic reflector, a small metal sheet reflector is sometimes used.

(ii) Rear Feed with Waveguide and Half-Wave Aerials

This is illustrated in Fig.822. The legs of the aerial are fixed on either side of a metal tongue projecting into the guide so that the electric field in the guide (H_{01} - mode) splits into two parts and excites the legs as it emerges from the guide.



(iii) Rear Feed with Waveguide and Slots

The guide terminates in a small cavity in which are two slots (Fig.823). These slots act as the radiating elements of the primary source.

Fig.822.- Rear feed with waveguide and half-wave aerials.

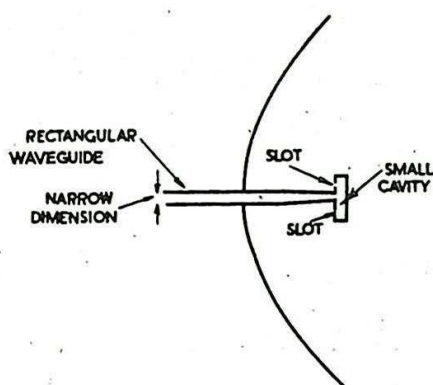


Fig.823.- Rear feed with waveguide and slots (long dimension of slots is perpendicular to the plane of the paper).

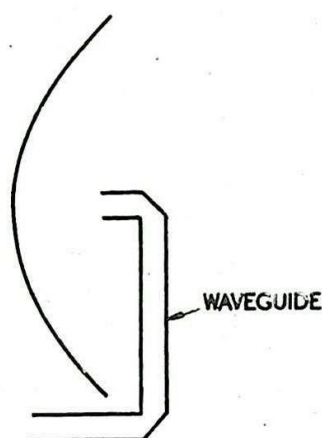


Fig.824.- Front feed with waveguide.

(iv) Front Feed with Waveguide (Fig.824)

The guide is brought round the edge of the mirror and the open end of the guide acts as the primary source. Little can be said about the field-strength distribution from the primary source at the open end of a guide. In general, the field-strength diagram is wide in a plane parallel to the narrow dimension of the guide and narrow in a plane parallel to the wide dimension.

In all the above types of primary feed, the phase centre is uncertain and must be determined by the designer in order that it may be placed at the focus of the mirror. The gains achieved in practice with these feeds are usually about 50% of the gain obtained with a uniform aperture distribution.

50. Beam Swinging with Paraboloidal Mirrors

The direction of the beam from a mirror can, of course, be altered by changing the direction of the axis of the mirror. This is not always feasible, however, and the alternative method of altering the phase distribution over the aperture is often used. If the aerial or other primary source is supported at the focus by means of an arm PF (Fig.825), and the arm is pivoted downwards through an angle θ to the position PF', the effect is to advance the phase of the radiation over the lower part of the aperture and to retard that over the upper part, so that the beam is tilted upwards by an angle α . For small angles of tilt the relation between θ and α is approximately

$$\alpha = 0.7\theta \dots\dots\dots (78)$$

It is not desirable to tilt the beam by an angle greater than about $\pm 5^\circ$ by this method, since the first side lobe then becomes too large.

51. Specially Shaped Mirrors

For some applications a narrow circular beam or pencil is not required but rather a beam which is narrow in the horizontal plane and

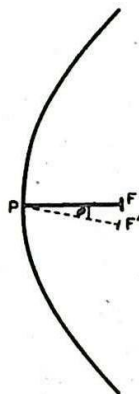


Fig.825.- Beam swinging.

specially fanned out in the vertical plane. Thus, suppose a radar set in an aircraft flying at height H is required to give equal illuminations on ships at different Ranges; (Fig.826). Now, due to the spreading out of the spherical wavefront from the transmitter the power density falls off as $1/R^2$ (see Sec.2). In order to compensate for this factor, therefore, one must arrange for the field distribution to be beamed more strongly in the direction of distant ships. From Fig.826 we have in triangle ABC

$$H/R = \sin \theta$$

$$\text{or } R = H/\sin \theta$$

$$= H \operatorname{cosec} \theta \dots\dots\dots (79)$$

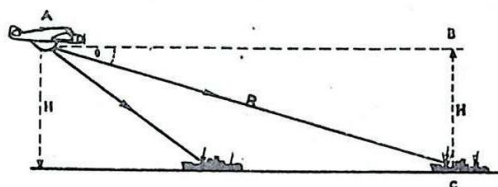


Fig.826.- Illumination of ships by aircraft.

Thus, if the power varies as $\operatorname{cosec}^2 \theta$, the factor $1/R^2$ will just be compensated and uniform illumination will be obtained. Power being proportional to the square of the field strength, the field-strength diagram of the aerial in the vertical plane must be of the form of $\operatorname{cosec} \theta$. To obtain a $\operatorname{cosec} \theta$ diagram the mirror must be suitably shaped. It appears that this shaping can be derived approximately from geometrical optics, i.e. assuming that the angle of incidence of each ray on the mirror is equal to the angle of reflection. From this it may be deduced that the mirror should take the form illustrated in Fig. 827.

Another method of shaping the radiation pattern in the required manner is to split the mirror in halves and move the upper section a little in front of the lower. The focus is not quite the same for each half but the difference may be neglected and a "mean focus" F assumed; (Fig.828). Consider two FAB and FCD, making equal angles with the

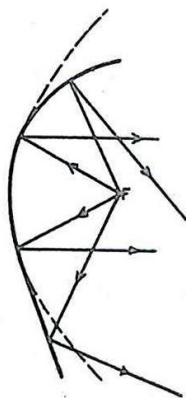


Fig.827.- Shaped mirror for $\operatorname{cosec} \theta$ beam.

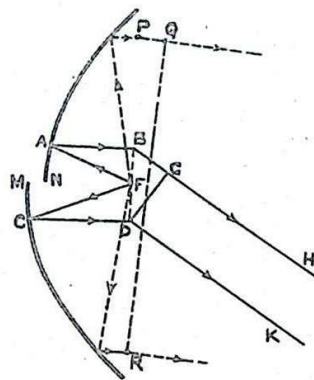


Fig.828.- Split mirror for $\operatorname{cosec} \theta$ beam.

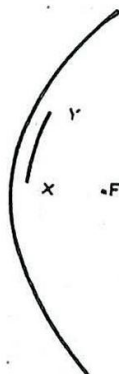


Fig.829.- Extra plate for $\operatorname{cosec} \theta$ beam.

axis. The ray FAB arrives at B in the aperture plane before FCD arrives at D. Now the aperture plane is considered as being the seat of a number of radiating sources. The source at B is leading in phase on the source at D. Thus these sources reinforce in the direction corresponding to the rays BH, DK, making an angle with the axial direction. The path difference between the rays BH, DK, is the distance BG, and this path difference must be just equal to the path difference between FAB and FCD.

This is approximately equal to $2MN$, where MN is the displacement between the two halves of the mirror. The downward deflection of the beam is more pronounced for pairs of rays near the centre of the mirror than for those near the edge. For example, two dotted rays are shown in Fig.828. In order that their path difference PQ should equal $2MN$ it is not necessary to deflect the outgoing rays very far downwards since these rays start from a very wide base line. In practice, therefore, only the centre portion is important for modifying the distribution in the required manner and, instead of the paraboloid being split, an additional metal plate XY is provided (Fig.829). This is attached by adjustable screws (not shown in the figure) and its distance from the main mirror is varied until the required distribution is achieved.

The gain of a cosec θ mirror is only of the order of half that of the undistorted mirror giving a pencil beam. Such a mirror has a field-strength diagram with no zeros. It was originally termed "An Ideal Gapless Radar Aerial Array".

52. Cheese Type of Mirror

The cosec θ aerial discussed in the last paragraph is elaborate and difficult to design. When it is merely desired to obtain a narrow beam in one plane (say the horizontal) and a less strictly specified wide beam in the other, the cheese reflector is probably simpler. It is illustrated in Fig.830, and consists of a metal parabolic cylinder with flat metal top and bottom plates. The aperture plane usually passes through the focus. The aim is to fill the aperture forming the mouth of the cheese with radiation and thus obtain a set of sources distributed over this area. The aperture then acts like a complete broadside array of rectangular shape and gives the appropriate field-strength diagram, narrow in the plane corresponding to the wide dimension and wide in the other plane. The primary feeding arrangements vary according to the polarisation of the wave in the guides.

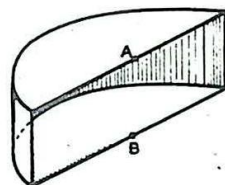


Fig.830.- Cheese reflector.

When the polarisation is such that the electric field (H_{01} - mode) is parallel to the top and bottom (or roof and floor) plates of the cheese, a flared rectangular waveguide is used. This is illustrated in Fig.831. The flaring takes place gradually in the wide dimension of the guide so that the H_{01} -mode is still present at the mouth of the flare. The width of the mouth is made $2/3$ the height of the cheese. It is shown in Chap. 5 that the distribution of electric field in the H_{01} -mode across

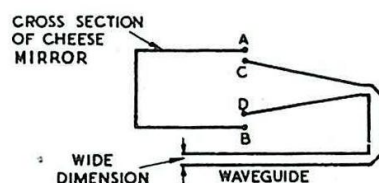


Fig.831.- Flared waveguide feed to cheese.

the wide dimension is sinusoidal in amplitude. This is shown in the full curve between the points C and D in Fig.832. The mouth of the guide is CD and the aperture of the mirror is AB , the distance between the roof and floor. The sinusoidal curve

between C and D can be resolved approximately into two sinusoidal curves fitted into the portion AB. We may regard the cheese mirror as a type of waveguide. There are two waves present in it, the dotted curve representing the amplitude distribution for an H_{01} -wave and the dot-dash curve representing that for an H_{03} -wave. The algebraic sum of these two waves gives the original H_{01} -wave emerging from the mouth CD of the flared guide. By the known properties of waveguides, these waves travel at different speeds. The depth of the cheese and its height are so adjusted that the two waves, after reflection at the curved surface of the cheese, arrive back in the plane of the opening in opposite phase. The resultant amplitude distribution is therefore now given by the difference between the dotted and dot-dash curves as shown in Fig. 833 (full curve). It is noticeable that the field is now distributed right across the aperture, but there is an undesirable dip in the centre. However, it appears that, probably owing to the presence of higher modes, the dip is much less pronounced in practice and a reasonably uniform distribution of amplitude is obtained across the opening of the cheese mirror in the short dimension.

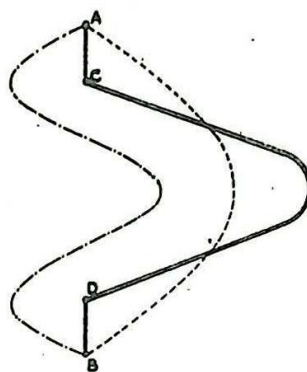


Fig. 832.- Distribution of field at mouth of flared guide.

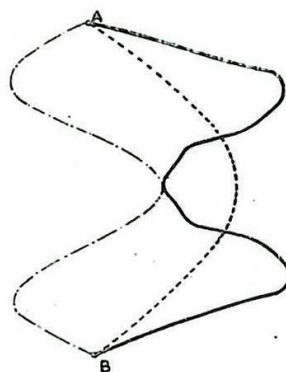


Fig. 833.- Distribution of field across narrow dimension of cheese aperture.

The distribution of amplitude across the long dimension depends on the field-strength diagram of the waveguide opening in this plane. By an argument similar to that in Sec. 48 it can be seen that, for uniform distribution across the opening, more power must be sent out from the guide in directions towards the edges of the cheese mirror, than towards the centre. This is unlikely to be achieved in practice, so that one can expect a tendency towards a tapered-distribution of power across the wide dimension of the aperture with corresponding diminution in gain, broadening of the beam and reduction of side lobes in comparison with the effects of uniform distribution across the aperture.

When the polarisation is such that the electric field is perpendicular to the upper and lower (or roof and floor) plates of the mirror, the method of flaring the waveguide feed does not apply. The difficulty now is to obtain a good distribution of power across the long dimension of the aperture since the waveguide is wide in this plane and directs the beam considerably into the centre of the cheese. Dielectric lenses, fastened to the waveguide mouth and designed to diverge the beam, have been used successfully.

53. Band-Width of Mirror-Type Aerials

In general, the frequency band over which mirror-type aerials will operate is much wider than that of broadside arrays. Consider first

the primary source, and suppose for example that this is the open end of a waveguide as in Fig. 824. In the absence of the mirror, the waveguide radiation into free space but there is some reflection at the open end of the guide. If a narrow slot is cut in the wide side of the guide and a probe inserted, a standing wave will be detected. The standing wave may be eliminated by the insertion of an iris or other device near the mouth of the guide. The standing wave ratio (SWR) is then 1:1. This matching by an iris is performed at a single frequency. It is found that the frequency can deviate by nearly 10% from this value without the SWR becoming worse than 1.2:1. Similar results are found with other types of feed.

Now let the mirror be placed in position with the primary source, matched to free space, at the focus. Some of the power after scattering at the mirror comes back and is picked up by the primary feed acting as a receiving aerial. A wave thus travels down the guide or feeder and a standing wave is created due to the presence of the mirror. If the primary source is matched to free space then the SWR when the mirror is in position is given by :-

$$S = 1 + \frac{g\lambda}{2\pi a}$$

where g is the gain of the primary source and a is the focal length of the mirror. The gain of most primary sources is about 6 or 7. As a numerical example take the case of a mirror with diameter 70 cm. and consequently with focal length $70/4$ or 17.5 cm. If the wavelength is 3 cm., then the SWR produced by the mirror is

$$1 + \frac{6.3}{2\pi \cdot 17.5}$$

$$\approx 1.2:1.$$

Smaller mirrors, with consequently smaller focal-lengths, produce a greater effect. The free-space matching is now altered so as to eliminate this standing wave. However, it is seen that the effect of the mirror is, in general, fairly small and depends linearly on the wavelength with a small factor of proportionality. The final result therefore is that a deviation of the order of several per cent can be made from the mid-frequency without undue standing waves being created on the feeder or waveguide.

The discussion above has centred round the question of the SWR on the transmission system. This is the important quantity in assessing the effect of the aerial on the transmitter valve (e.g. magnetron). Provided the SWR is not more than about 1.5:1, reasonably efficient operation is to be expected.

The alteration of the SWR due to the introduction of the primary feed into the mirror can be eliminated. This is done by placing a small raised metal plate at the apex of the mirror of such size and such position that the power scattered back to the focus is zero owing to interference effects from different parts of the mirror and plate. This gives a broader band of operation. Unfortunately this modification increases the side lobes and reduces the gain.

54. Slots in Waveguides

Consider a rectangular waveguide carrying an H_{01} -wave.

Electric currents flow on the inside walls of the guide as shown in Fig. 824. Essentially there are two directions of current flow, along the centre of the wide walls of the guide and transversely across the narrow walls.

The longitudinal currents may be regarded as carrying the power along the guide and the transverse currents as shunt currents which help to sustain the wave pattern but do not normally take part in the flow of power. If a thin slot AB is cut along the centre of the wide wall, the effect on the currents is negligible and power continues to flow. Similarly a thin transverse slot CD across the narrow wall does not affect the flow. On the other hand a transverse cut EF across the wide wall interrupts the currents, and radiation of power takes place from this slot. One may think of the transverse magnetic field in the wave pattern arriving opposite the slot and some of this magnetic field escaping. The magnetic field in the radiated wave is parallel to the length of the slot. Because it interrupts the longitudinal currents in the guide, the slot EF is regarded as series connected; (see also Chap. 5 Sec. 40).

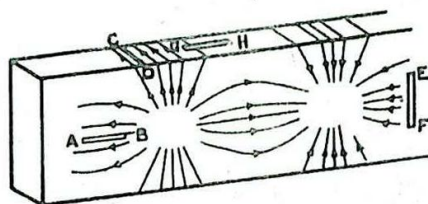


Fig.834.- Currents in walls of rectangular guide.

A longitudinal cut GH on the narrow side of the guide interrupts the shunt currents. It radiates in a manner similar to EF, being excited by the longitudinal magnetic field in the wave pattern. It is regarded as shunt-connected.

If only currents in the direction of propagation are interrupted then the slot is series connected; if only currents perpendicular to the direction of propagation are interrupted then it is shunt connected.

Slots such as EF and GH would radiate very fiercely, that is to say, they are tightly coupled to the guide and have, respectively, a high resistance and high conductance. The coupling may be weakened by altering either the position or the inclination of the slot. Positions 1, 2 and 3 in Fig.835 indicate successive loosenings of the coupling for a shunt slot. Fig.836 shows the same thing for alteration in inclination.

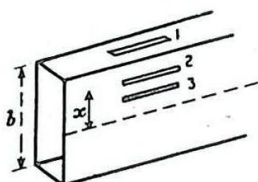


Fig.835.- Shunt slot in guide - different positions.

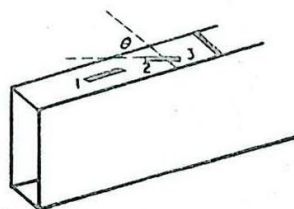


Fig.836.- Shunt slot in guide - different orientations.

Quantitative results are available for some shunt slots. Such slots are regarded as shunt elements connected across a transmission line whose characteristic impedance is taken as unity. The normalised conductance thereby obtained is then a measure of the power radiated.

In Fig.835, if x is the displacement from the centre line and b is the width of the broad side of the guide, the conductance is proportional to $\sin^2 (180^\circ x/b)$. For a 3" guide with slot 2" long and 3/16"

wide used for wavelengths of about 10 cm., the factor of proportionality is about $\frac{1}{2}$, so that the maximum effect, when $x = \frac{b}{2}$ is that of an element

of resistance 2 ohms across a line whose characteristic impedance is 1 ohm. Such a slot will radiate $\frac{1}{3}$ of the power in a wave travelling in the guide, the remaining $\frac{2}{3}$ passing on. In Fig. 836, if θ is the angle of rotation of the slot from the non-radiating position, the conductance is proportional to $\sin^2\theta$. For a guide $\frac{1}{2}$ "x1" outside dimensions, of 18 gauge wall, with slot 0.14" deep and 0.0625" wide used for wavelengths of about 3 cm., the factor of proportionality is about $\frac{3}{4}$.

55. Slotted Linear Arrays

Based on the theory of the last section various kinds of linear arrays have been constructed with waveguides. A simple example is shown in Fig. 837 with shunt slots, about a half-wavelength long, spaced half a guide-wavelength λ_g along the waveguide. The end of the waveguide is closed with a metal plate $\lambda_g/4$ from the centre of the last slot. Successive slots are on opposite sides of the centre line in order to compensate for the reversal of phase in the standing wave inside the waveguide. In designing this type of array the displacement of slots from the centre line is chosen so that on the whole the guide is roughly matched, no power being reflected back. Variants of this array have the slots on the centre line and screws at the edge of each slot to distort the field inside the guide, causing the slot to radiate. Generally speaking, screws are troublesome due to their tendency to spark.

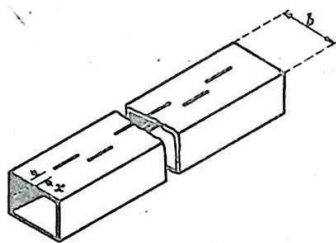


Fig. 837.- Slot array with transverse polarisation.

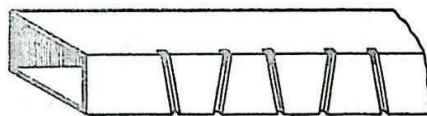


Fig. 838.- Arrangements of slots in waveguides.

The type of slotted array in most favour at the moment employs cuts in the narrow side of the guide (Fig. 838) inclined so as to give the required coupling. These are again shunt slots. At the same time a non-resonant design is adopted. The waveguide array as described earlier in this paragraph has a metal piston at the end and there is a substantial standing wave inside the guide. Such a system is said to be resonant and suffers from the defect that when the frequency deviates slightly from its correct value, the matching and the radiation pattern deteriorate appreciably. For a 50-element array the overall frequency bandwidth is 1% of the mid-frequency; for 200 elements it is $\frac{1}{2}\%$. In the non-resonant design an attempt is made to produce a single travelling wave along the guide and no reflected wave. This is effected by making the centre-to-centre spacing of the slots a little different from $\lambda/2$ (e.g. 200° of phase instead of 180°). Then waves reflected at the slots do not reinforce one another to create a standing wave but rather tend to cancel one another out. The coupling of the slots is gradually tightened as the distance along the guide from the input end is increased in such a manner that all the slots radiate the same power. The small amount of power left in the guide after the last slot is often absorbed in a non-reflecting dummy load.

The following points are of interest regarding non-resonant arrays with slots on the narrow side :-

- (i) There is unwanted polarisation in the beam owing to the inclinations of the slots. This is said to be small in practical installations.
- (ii) The slots have to be cut carefully at the correct angle, and the fact that the angle is different for every slot introduces manufacturing difficulties. The difficulty is overcome to some extent by allowing the slots to be arranged in a succession of small groups with all the slots in a group inclined at the same angle.
- (iii) Since the slots are not being driven in phase (not being spaced at intervals of $\lambda/2$) the beam is swung away from the normal to the array (see Sec. 38); in fact the beam forms part of a cone with the array as axis: this is sometimes undesirable and is eliminated by adding two triangular-shaped metal sheets (Fig. 839) which confine the waves radiated from the slots for appropriate distances so that they emerge in phase into space.

56. Slotted Waveguide with Mirror

The rays emerging from the slotted waveguide array form a narrow beam radiating approximately at right angles to the array, and all the considerations of Sec. 30 and Fig. 786 apply. In the plane at right angles to the array the beam is very broad. By directing the radiation from the waveguide into a mirror (e.g. a parabolic cylinder of about the same length as the array (Fig. 840)) the original beaming is retained and, due to the mirror, beaming is also introduced in the other plane.

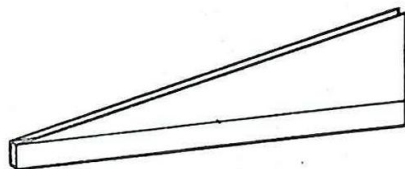


Fig. 839.- Addition of triangular plates to slotted waveguide array.

57. Microwave Beacon Aerials

As remarked at the end of Sec. 30, a beacon aerial is usually required to radiate all round but with a given vertical coverage. An early type of microwave beacon aerial was the Biconical Horn (Fig. 841) excited by a small vertical aerial at its apex. Vertical linear arrays are, however, coming into use. For horizontal polarisation groups of three half-wave centred aerials are placed every half-wavelength up the length of a coaxial feeder. Probes protruding through holes in the outer of the coaxial line are used to excite the aerials (Fig. 842). The aerials stand off about quarter of a wavelength from the coaxial. Provided the point A in Fig. 842 is at a low voltage, the quarter-wave stub BAC tends to make the voltages at the points B and C oscillate antiphase. Thus the point C tends to be excited by the energy in the probe. The point D is connected to C. Thus, D and B are excited antiphase, power travels out between BA and DF and the left and right hand parts of the aerial are energised. Similar considerations apply to the other aerials. The resultant field-strength distribution is almost uniform in a horizontal plane.

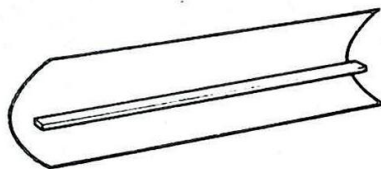


Fig. 840.- Waveguide array feeding parabolic cylinder.

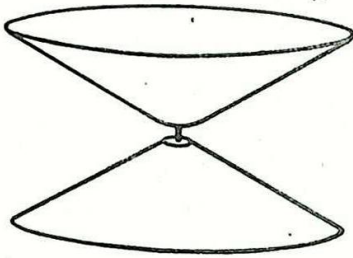


Fig.841.- Biconal horn.

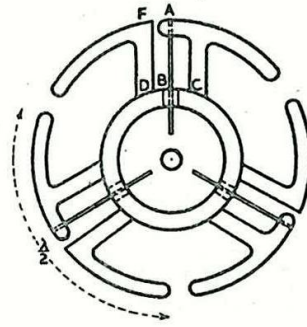


Fig.842.- Aerial array giving horizontally polarised field and a uniform field strength distribution in a horizontal plane.

58. Effect of Ground on Microwave Aerials

In Sec. 46 it was shown that the radiation from an array of aerials is affected by proximity to ground. A mirror which gives horizontally polarised waves can be replaced by an equivalent array of resonant aerials; hence its associated image can be deduced; (Fig.843). In some applications a cheese mirror (not elevatable) is used, which gives a field-strength diagram (neglecting the effect of ground) symmetrical above and below the horizontal. Let the corresponding beaming factor be denoted by $B(\alpha)$. Such a system can be replaced by isotropic sources at the aerial and image phase centres, and the effect of the ground deduced from the interference pattern of the two sources corresponding to a field strength factor $G(\alpha)$. For such a mirror the results of Sec. 26 would hold. If the system is many wavelengths above ground there will be a large number of maxima and minima in the vertical field-strength diagram.

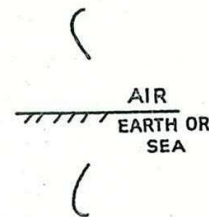


Fig.843.- Image of mirror aerial with horizontal polarisation.

59. CONCLUSION

In this chapter there has been an attempt to describe a number of types of aerial systems, and in addition to formulate the general principles of transmission and reception. Emphasis has been laid on the broadside array and its centimetre-wave counterpart, the paraboloidal mirror; also on the end-fire aerial, including the Yagi. There are, however, other aerials to be found in use with pulse systems and some of these are now mentioned briefly :-

The V-reflector aerial is one in which one uses a V-shaped instead of a parabolic reflector. The area of the aperture of the mouth of the V and the efficiency with which it is filled with radiation from the primary source are the factors determining the performance.

The polyrod or dielectric aerial is coming into use. It is an end-fire type of aerial formed by a dielectric rod along which power

travels, as in a waveguide, but at the same time a fraction of the power leaking out into space all along the surface of the rod.

The slot aerial is sometimes used as a single element. Slots in waveguides have already been discussed in Secs. 54 and 55. A single $\lambda/2$ slot appropriately placed in the wall of any kind of cavity will act as a radiator, the plane of the magnetic field in the radiated wave being parallel to the length of the slot.

The long-wire travelling-wave aerial, sometimes used on metre wavelengths, is already familiar, being described in detail in standard works on radio.

CHAPTER 18

CONTROL SYSTEMS

1. INTRODUCTION

A control system, in general terms, consists of an arrangement of elements (amplifiers, converters, human operators, etc.) interconnected in such a way that the operation of each depends on the results of the operation of one or more other elements, and the purpose of which is to control some process or machine.

The whole system commences with the input element, the operation of which should be independent of the control system, and the output element is the one which affects directly the process or machine to be controlled. There may be several input or output elements. In most cases in radar the output element controls the position of the load, which may consist of a needle indicator, a shaft leading to some other control system, an aerial array or cabin which must be rotated to follow a target, or some other device.

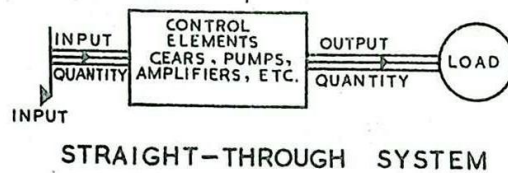
A common example of a control system is the accelerator of an automobile engine; the position of the pedal, or input element, controls the applied engine torque which, together with external conditions depending on the resistance to motion, inertia, etc., determines the speed of the car. The road wheels may be regarded as the output elements. Such a system could be described as a position-velocity system, or more exactly, a displacement-controlling-velocity system, since the position or displacement of the input element ultimately controls the velocity of the output elements. It should be noted that the relation is not necessarily a linear one, and that it will be affected by differing external conditions.

The above is an example of an automatic, power-amplifying control system. In the case of a recording barometer, the variation in air pressure acting on a diaphragm usually provides the power to operate the recording device directly, so that the system is not power-amplifying. The supply of air to a pipe organ, on the other hand, requires power to be supplied by a motor or by a human operator; in either case power amplification is provided, a pressure gauge indicating the quantity of air required to maintain an adequate bellows pressure, and the motor or organ blower responding accordingly.

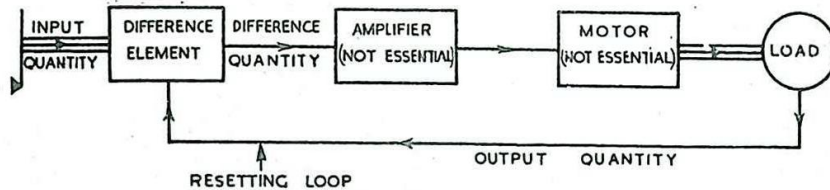
The mechanism of an automatic control system may be hydraulic, mechanical or electrical. In any case the essential links are indicated in Fig. 844. Fig. 844(a) shows a "straight through" arrangement in which input controls output directly, as with a Bowden cable (non-power-amplifying) or as in the motor-car throttle control (power-amplifying).

The first part of this chapter is devoted to a consideration of electrically operated data transmission systems, which may be represented schematically by the arrangement of Fig. 844(a), particular emphasis being given to remote position-indicating devices. The second part deals with Servo Systems.

A servo system is a control system which is Error-Actuated. As indicated at (b), at some stage of the control sequence the output quantity, or some function of it, is compared with a similar function of the input quantity, so that the whole system operates on the principle of reducing to zero the error, or difference between input and output. The process of comparing output with input is called Resetting. A servo is an automatic power-amplifying reset control system.



(a)



(b)

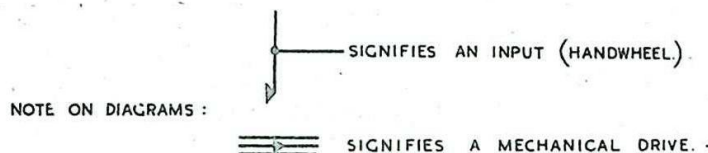


Fig. 844 - Fundamental arrangements of elementary control systems.

According to this definition, certain electronic circuits are servos whereas others are not. A simple valve amplifier is not a servo, since there is no reset. Some feedback amplifiers, on the other hand, employ the resetting principle and thus come into the servo category. For example, in a cathode follower the output quantity (cathode potential) is compared with the input quantity (grid potential) and the power amplifying properties of the valve are utilised to minimise variations in the difference between the two (error voltage).

DATA TRANSMISSION SYSTEMS (NON-RESETTING)

2. General

The function of a Data Transmission System is to convey to some remote receiver data which is represented by conditions at a transmitter. Usually this data is represented both at the transmitter and at the receiver by a mechanical movement, but this is not necessary. Such systems may be variable either smoothly or step-by-step; the transition from the one type to the other is gradual and ill-defined. For example, a wire-wound potentiometer is essentially a step-by-step device, since the division of the total resistance into its two parts changes with movement of the potentiometer arm by finite steps, corresponding to the resistance of a single turn of the wire. In effect, however, the magnitude of these steps can be so reduced as to make the error negligible if the device is assumed to be smoothly variable.

The commonest types of data requiring transmission in radar are those giving numerical indication of the position of the transmitter; e.g. the Bearing or Elevation of an aerial system. Similarly, measured Range may be indicated by the angular movement of a calibrated transmitting shaft, although it may be transmitted as a variable voltage, the indication at the receiver being by means of a calibrated voltmeter.

A data transmission system is Synchronous if there is negligible delay between the setting of the transmitter and the adaptation of the receiver in conformity with the conditions at the transmitter; i.e. the follow-up time may be taken as zero. Systems which embody servo action usually exhibit a time-delay in operation which may or may not be sensibly constant for all inputs. They can seldom be considered as synchronous.

A system is linear if equal movements at the transmitter give rise to equal changes in indication at the receiver. Where duplicate systems are used, transmitting coarse and fine data, it is usually essential that the systems should be linear, otherwise one revolution of the fine transmitter would not correspond to the same incremental movement of the coarse one. The alternative, a non-linear gearing system, is not normally practicable. The use of such coarse-and-fine systems makes great accuracy possible without complicated or finely machined construction. For example, in a rotational system a maximum error of one part in 3,600 means, with a single, or coarse, indicator, an accuracy of $\pm 0.05^\circ$ in 360° . With a coarse-and-fine system an error of $\pm 3^\circ$ can be permitted in both indicators without any deterioration in accuracy. This allows greatly increased manufacturing tolerances and often increases the ease of operation.

Various non-electrical synchronous transmission systems are in everyday use, but only electrical methods are dealt with in the following sections. Mechanical devices, such as flexible-cable rotary drives or push-pull controls can be used, but are inconvenient over distances of more than a few yards and in any case are prone to fatigue. Hydraulic and pneumatic devices have not been extensively applied to service radar transmission problems. Electrical methods require only a multi-core cable between transmitter and receiver, and great accuracies may be maintained over long distances.

3. Performance of Transmission Systems

The criteria by which a transmission system is judged depend largely upon the use to which it is put. If it is required to operate a light pointer only, power considerations generally, and efficiency in particular, are not usually important. If it is required to drive a heavy mechanical load these considerations may predominate over all others. If the system is not power-amplifying, the power to drive the load at the receiver must be provided by the source operating the transmitter. Systems which are power-amplifying are frequently of the servo type, the transmission being carried out at a low power level and the signal power being amplified at the receiver by a servo.

The criteria usually applied to transmission system are as follows:

- (1) Stiffness: this is defined as the magnitude of the disturbing torque that has to be applied to the load to displace it by a unit angular amount from its position corresponding to a fixed input.
- (ii) Maximum static errors with or without load.

- (iii) Dynamic errors; velocity and acceleration lags.
- (iv) Frequency response.
- (v) General reliability, size, weight, etc.

Some of these criteria are dealt with more fully in Secs. 15 and 17-19.

4. Step-by-Step Systems for Transmitting Position Data

(i) Lamp system. In this system a lamp lights at the receiver to indicate the position of the moving arm at the transmitter. The very simplicity of this system, as indicated by the circuit of Fig. 845, makes it particularly useful where reliability is of paramount importance. The method can be adapted to any degree of accuracy by the use of coarse and fine dials with lamps at regular intervals on the circumference. Even so, a large number of cable cores is required. Since this is an On-Off system no errors are introduced

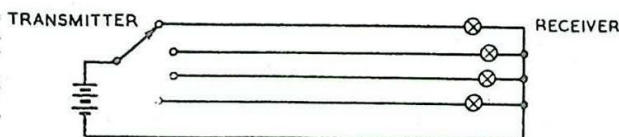


Fig. 845 - Lamp system.

by line losses, provided the lamps always light when (and only when) they are switched on. By the use of gas-filled lamps the current can be kept very small. Whilst the method is adaptable to pointer-matching receivers it is purely an indicating system and is not readily adapted for power drives. Either AC or DC supplies may be used.

(ii) M-Type System. The M-Type Transmission System, or M-Motor as it is sometimes called, is a relatively high-power analogue of the lamp system, the receiver current being used to provide a driving torque instead of energising a lamp. The method is illustrated in schematic form in Fig. 846.

As the brush at the transmitter rotates and "makes" on each segment in turn of the transmitter switch, the corresponding armature winding at the receiver is energised, and the magnetic field set up produces a torque on the soft-iron rotor, which is magnetically asymmetrical, tending to pull its magnetic axis into line with the field. If the brush makes on two segments at once, both of the corresponding windings are energised and the rotor tends to take up a position intermediate between the field

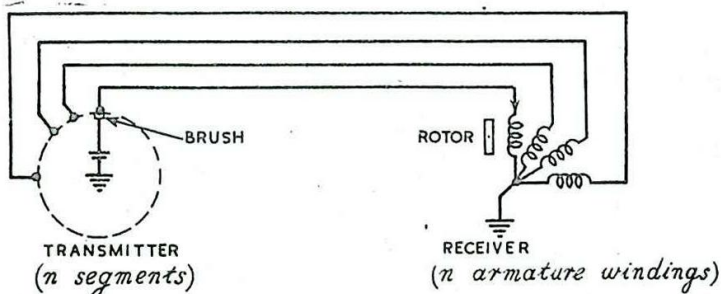


Fig. 846 - Schematic arrangement of M-type transmission system.

axes due to the individual windings. In practice this overlap interval is often brief and then does not greatly reduce the inherent errors of the system. If there are n segments the maximum off-load error (i.e. assuming there is no load torque) is $\frac{360^\circ}{2n}$. This possibility of error is accompanied by a 180° ambiguity if an unpolarised rotor is used, since in this case a reversal of the rotor does not change the magnetic stability of the system. If the rotor is polarised, two positions of equilibrium remain for each position of the transmitter brush, but one of these is unstable and of little practical importance. As the transmitter arm is rotated steadily the receiver rotor follows in jumps of $\frac{360^\circ}{n}$. The relative dispositions of brush and armature windings should normally be adjusted so that the error is zero when the brush is in the centre of each segment.

A variant of the elementary circuit is shown in Fig. 847. By the use of three fixed brushes and dead segments on the rotary switch the field at the receiver is made to rotate in steps of 30° , corresponding to a 12-segment switch of the elementary type shown in Fig. 846.

The receiver is shown wound for a two-pole field. The use of a four pole winding in the receiver reduces the step to 15° , but at the expense of introducing 180° ambiguity, even with a polarised rotor. Such a four-pole winding is illustrated in Fig. 848. Only one complete winding is shown, for the sake of clarity. The windings for the sets "2" and "3" are the same as for "1", each starting at the appropriately numbered point and ending at an earthed point. If the transmitter of Fig. 847 is used with such a receiver there is a 1:2 ratio between the movement of the rotor and that of the transmitter switch.

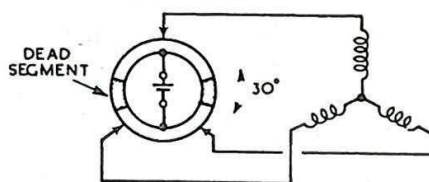


Fig. 847 - M-type transmission.

Fig. 849 shows an alternative form of transmitter switch providing a 1:1 ratio when used with a four-pole receiver. Neglecting the brief overlap period when a brush bridges two segments, the receiver field advances in 30° steps, i.e. 12 per revolution. There is a 180° ambiguity even with a polarised rotor.

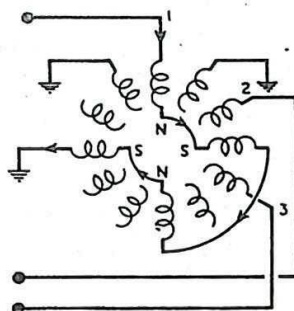


Fig. 848 - 4-pole receiver.

A transmitter switch which provides 24 steps/rev. is shown in Fig. 850. Segments similarly numbered are connected together. The effect if this is used with the receiver of Fig. 846 is the same as if a 1:2 (step-up) gear were introduced between transmitter and receiver. The brushes may be duplicated, if desired, as shown (unshaded). If an 8-pole receiver is used, the 1:2 ratio is removed, but there are four stable positions of equilibrium with a polarised and eight with an unpolarised rotor.

The principal advantage of the M-type system is that it combines power amplification with relatively high efficiency. It may be used with an alternating supply provided the rotor, if polarised, is fed from an inphase source. Its chief disadvantage lies in the large number of ambiguities which must be permitted if small incremental steps are to be obtained.

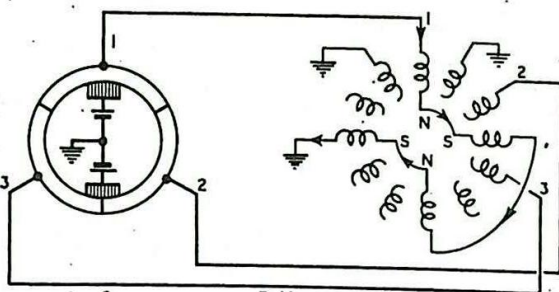


Fig. 849 - Alternative form of M-type system.

In its practical form the M-motor is not Self-Aligning; i.e., once the receiver is out-of-step with the transmitter it requires some external cause to bring it into step again, since any one of the ambiguous positions of the receiver rotor is as stable as any other.

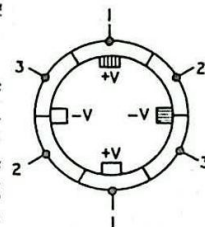


Fig. 850 - Transmitter providing 24 steps/rev.

5. Potentiometer and Voltmeter Systems

The basis of many transmission techniques is the potentiometer and voltmeter arrangement of Fig. 851. The voltmeter, in the simple circuit, is calibrated to indicate the displacement of the potentiometer slider from the zero end. Theoretically, if a wire-wound potentiometer is used, the indications are stepped, but in practice the steps may be smaller than the static error of the meter. Dynamic accuracy depends on the stiffness/inertia and damping ratios of the meter; (see Sec. 19). The accuracy is affected by anything causing variation in load current, such as line resistance or fluctuations in battery voltage. Changes of temperature may affect the accuracy of the voltmeter through mechanical changes or through variation of resistance.

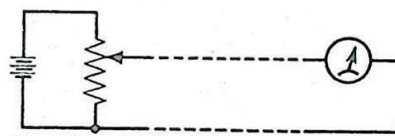


Fig. 851 - Potentiometer and voltmeter system.

All of the systems employing the basic principles of the circuit of Fig. 851, some of which are described in Sec. 6, are essentially power-amplifying. That is to say, the power which provides the output torque to bring the needle or rotor into alignment may be much greater than the power necessary to move the input potentiometer arm. On the other hand, these systems are not suitable for high powered drives, because of their very low efficiency. In general the load current must be small compared with the current flowing in the various branches of the transmitter potentiometer, so that if high powers were required at

the receiver prohibitive dissipation would be necessitated at the transmitter. In this sense these systems compare unfavourably with the non-power-amplifying selsyns, described below, or the power-amplifying M-motor, (sec. 4).

The simple circuit of Fig. 851 may be modified by the substitution of an AC supply and meter for the DC arrangement. This introduces further sources of inaccuracy, of which variations in waveform and frequency of the supply are the most prominent.

One form of voltmeter commonly used is indicated in Fig. 852. This requires a 3-core cable, but is independent of the supply voltage. Such an arrangement is fundamental in all systems which are independent of supply fluctuations, and, in some cases, of line resistance (provided this is the same for all the lines).

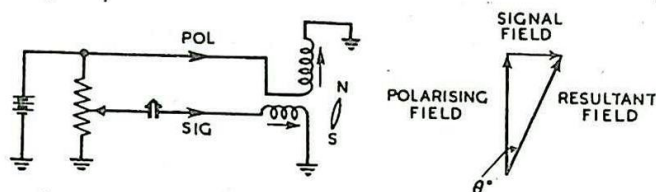


Fig. 852 - Electrodynamic system with DC polarized voltmeter.

The direction of the resultant field at the receiver depends on the position of the slider at the transmitter, and is indicated by a magnetised rotor. This rotor aligns itself with the resultant flux, which is inclined at an angle θ_0 to the flux axis of the polarising winding. The system is non-linear, the movement of the input slider being proportional to $\tan \theta_0$. For DC operation a permanently magnetised needle may be used. For AC working the needle must be magnetised from a source in synchronism with the potentiometer supply.

6. Ring-Potentiometer Systems

The resolvers described in Chap. 3 Secs. 18-19, may be adapted as transmitters or receivers for use in data transmission systems, as indicated in Fig. 853. In this case sine-graded resistive potentiometers are used to form the transmitter, fed from a DC source, and a pair of inductive resolvers with a single magnetic rotor, which aligns itself with the resultant flux, form the receiver. The transmitter potentiometers are so wound that

$$V_x \propto \sin \theta_1$$

and $V_y \propto \cos \theta_1$, where θ_1 is the angular rotation of the mutually perpendicular potentiometer arms from some reference position. Provided each resolver is wound so as to produce a uniform flux of magnitude proportional to the input voltage, the inclination θ_0 of the resultant field at the receiver, is given by

$$\tan \theta_0 = \frac{V_x}{V_y} = \tan \theta_1$$

$$\text{Hence } \theta_0 = \theta_1 + n.180^\circ$$

This system is therefore linear. The ambiguity ($n = 0$ or 1) may be resolved by the use of a polarised rotor, in which case one of the equilibrium positions is unstable. If an unpolarised rotor is used the receiver must have identical scales, each occupying 180° , if ambiguity is to be avoided; (Fig. 854).

The system may be used with an alternating supply. In this case inductive potentiometers may replace the resistive ones. The rotor must be polarised from an AC source in synchronism with that which feeds the potentiometers.

An alternative arrangement of the sine-graded ring potentiometer system is shown in Fig. 855. This is typified by the term Inverted, indicating that the supply is fed to the potentiometers via the rotatable potentiometer arms, the outputs to the receiver being taken from fixed points. Separate ring potentiometers are necessary for the x and y windings of the receiver.

Although the use of sine-graded potentiometers gives a theoretically linear relation between θ_1 and θ_0 , whereas non-sinusoidal potentiometers introduce non-linearity, the latter are commonly used in practice because of their comparative simplicity of construction. By the adoption of coarse-and-fine methods the errors introduced by the employment of uniformly-wound potentiometers may be made of negligible importance. Such a system is indicated in Fig. 856 (a) and is analogous to the phase-shifting network described in Chap. 3 Sec. 23. The errors introduced by the potentiometer linearity are the same as for the phase-shifting network and the cause is illustrated in Fig. 856(b). This is a vector diagram indicating the direction $O'P'$ of the resultant flux in the receiver. Using the symbols as indicated, to represent component fluxes and voltages, we have $B_y \propto V_y$ and $B_x \propto V_x$, the changes in each quantity being proportional to the angular movement θ_1 of the potentiometer arm. As P moves from A to B (Fig. 856 (a)) P' moves from A' to B' (Fig. 856(b)), the distance $A'P'$ being proportional to the change in voltage from A to P, i.e. proportional to θ_1 . If θ_1 is measured in degrees, and $\phi = \frac{\theta_1}{90}$,

then $B_x \propto \phi$

and $B_y \propto 1 - \phi$

so that $\tan \theta_0 = \frac{B_x}{B_y} = \frac{\phi}{1 - \phi}$.

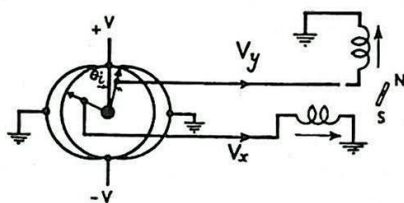


Fig. 853 - Two-coil receiver fed from sine-graded potentiometer.

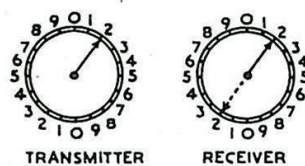


Fig. 854 - Resolution of ambiguity due to unpolarised rotor.

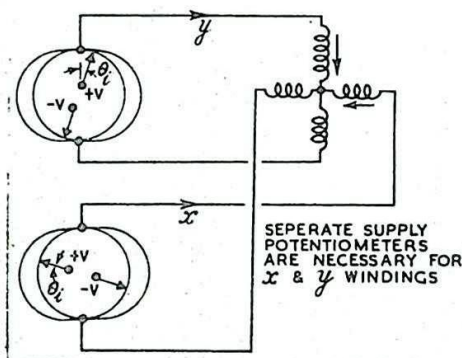


Fig. 855 - Inverted sine-graded transmitter.

The output alternately lags and leads the input. The error $\theta_o - \theta_i$ is zero when $\theta_i = 0^\circ, 45^\circ, 90^\circ$ etc. Maximum error occurs when $\theta_i = \pm 21.5^\circ, 90^\circ \pm 21.5^\circ$, etc; its value is $\pm 4.1^\circ$.

Since the potentiometer is uniform it may be used either as shown in Fig. 856 (a) or inverted, as in Fig. 857. This latter circuit also uses a full-wave or balanced arrangement of feed-points and receiver coils. An advantage of this arrangement is that the accuracy is not affected by voltage drops at the potentiometer output connections due to load current, provided the load is balanced and symmetrical; i.e., the input resistance of all the stator coils, measured between each input terminal and earth, is the same, and the four cable connections have equal resistance. The off-load accuracy (i.e. ignoring the effect of current drain on the transmitter) is the same for this arrangement as for that of Fig. 856 (a).

A modification of the method of Fig. 857 is shown in Fig. 858 where a three-coil receiver is fed from an inverted linear ring-potentiometer. This results in an improvement in off-load accuracy, the maximum error being reduced to 1.1° . The relative voltages carried by the three lines are indicated in the figure in terms of ϕ , where $\phi = \frac{\theta_i}{90^\circ}$.

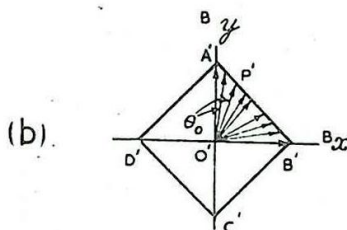
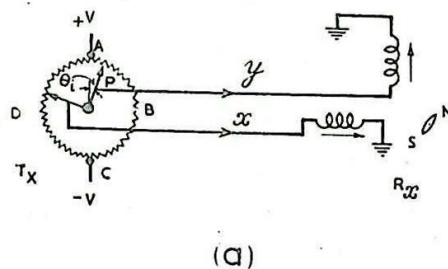


Fig. 856 - Uniform ring-potentiometer system.

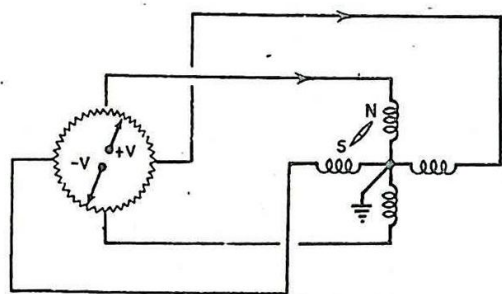


Fig. 857 - Inverted full-wave arrangement.

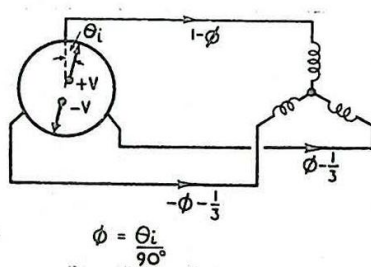


Fig. 858 - Three-coil receiver fed from linear ring-potentiometer.

It may be deduced that, with the same symbols as for Fig.

$$B_y \propto \frac{4}{3} - \phi$$

and $B_x \propto \sqrt{3} \phi$;

$$\text{so that } \tan \theta_0 = \frac{B_x}{B_y} = \frac{\sqrt{3}\phi}{\frac{4}{3} - \phi} = \frac{\sqrt{3}}{\frac{120}{\theta_1} - 1}$$

Provided a polarised rotor is used, no error is introduced at $\theta_1 = 0^\circ$, 30° , 60° etc. Maximum error occurs at $\theta_1 = \pm 13.2^\circ$, $60^\circ \pm 13.2^\circ$, etc.

7. Single-Phase AC Systems

Any of the systems described in Secs. 1 - 6 may be operated from single phase supplies, although generally speaking the accuracy is not as good as with DC. The systems dealt with in the following sections cannot be operated from DC supplies since, in all cases, the transmitter EMF's are magnetically induced, and, in some cases, the receivers operate correctly only if the frequency of the alternating supply is maintained at a suitable value.

The term Selsyn (from Self-Synchronous) is used for either a transmitter or a receiver designed on the principle of the inductive resolver or variable-ratio transformer. (Chap. 3 Sec. 19). The

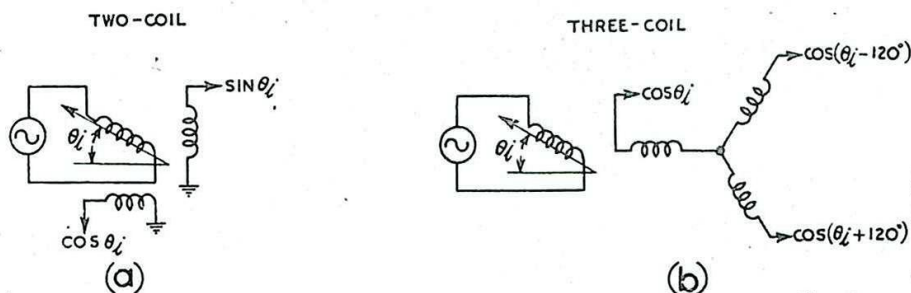


Fig. 859 - Selsyn transmitters.

transmitter fulfils the function of the sine-graded potentiometers in the circuit of Fig. 853, feeding to each receiver coil a voltage proportional to the cosine of the angle between the transmitter rotor coil and the appropriate stator coil. Fig. 859(a) shows schematically a two-coil stator and Fig. 859 (b) a three-coil stator selsyn transmitter. The currents in the receiver coils have the same function as in the DC circuit, to produce a resultant flux with which the polarised rotor aligns itself. In the selsyn receiver, as in the selsyn transmitter, the soft-iron core of the rotor is polarised from an AC supply fed to the rotor winding via slip-rings.

In the Magalip (Magnetic Slip-Ring) receiver a shaped soft-iron rotor is polarized by magnetic induction from a stationary winding so that the use of slip-rings is avoided. This results in a very considerable loss in efficiency, the output torque available from a magalip receiver being sufficient to drive only a light pointer.

(Note: The term Magalip, which originally had only the meaning assigned to it above, is now employed to describe many other transmission components, some of which do not employ slip-rings, and is applied to some devices in which no movement at all occurs. The transmitters used with magalip receivers are called magalip transmitters or transmitter magalips, although they are designed on the selsyn principle, using slip-rings).

The advantages afforded by the use of selsyn-type transmitters instead of resistive potentiometers are :-

- (i) they are more compact for a given output power;
- (ii) they are easier to make to a given degree of accuracy;
- (iii) they have a lower output impedance for a given input power, so that there is less reaction due to loading by the receiver. Also, a slightly unbalanced load does not seriously unbalance the output voltages of the transmitter.

A disadvantage is that selsyn systems are not power-amplifying. Because of its high efficiency a "power" selsyn (i.e., a large one) may be employed to drive a load which requires a large driving torque, but this torque must be provided by whatever source of power is used for driving the rotor of the transmitter. In this sense selsyns are inferior to M-motors, but compared with the latter selsyns have the advantages of :-

- (i) smoothly variable output, instead of step-by-step;
- (ii) absence of ambiguity; (i.e., they are self-aligning);
- (iii) they require a cable with only five cores, whereas an M-motor may need many more.

8. Selsyn Systems

Fig. 860 shows a simple two-circuit selsyn system. Such a system is sometimes described as a "Two-Phase" system. The term is not used here, since "phase" is reserved for time-variations. Many of the windings used in single-phase AC transmission systems are similar to those found in polyphase motors or generators, where a rotating flux is generated. In the systems here described single phase supplies are used throughout. A flux can be made to rotate only by rotating the

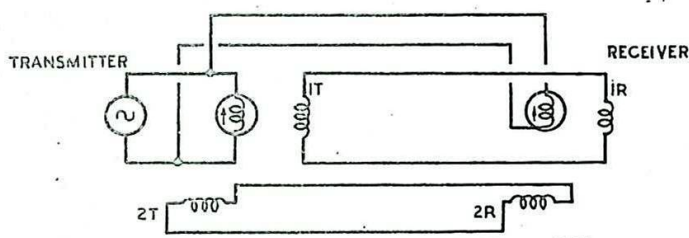


Fig. 860 - Simple two-circuit selsyn system.

appropriate field windings.

In the arrangement of Fig. 860 the rotor of the receiver is polarised from the same source as feeds the transmitter. Transmitter and receiver may be identical units, or the receiver can be of lighter construction. In the former case, when the receiver is in alignment with the transmitter the line current is zero, since the induced EMF's cancel in each of the secondary circuits. Each of the primary windings (the two rotors) draws from the supply sufficient magnetising current to set up a flux which gives a primary back EMF equal to the excess of the supply voltage over the series resistance drop. The arrows in the diagram show the flux directions for a chosen half-cycle of the supply current. Fig. 861 shows a three-circuit system. A three-circuit selsyn is preferable to a two-circuit one because :-

- (i) The problem of winding the secondary circuits is simplified.
- (ii) Even if the variation of mutual inductance with rotor angle is not truly sinusoidal, the error introduced at the receiver is zero every 30° , compared with every 45° for the two-circuit system; and the maximum error is less.

The three-circuit selsyn system does not require any more cable cores than the two-circuit system.

In general several receivers may be driven from a single transmitter. Where receivers and transmitters are of identical construction interaction between receivers may be prohibitive. If one receiver is misaligned currents flow from the other receivers providing a correcting torque, and these currents cause the other receivers to be misaligned also. This interaction may be reduced by making the receivers of high input impedance compared with the output impedance of the transmitter, so that misalignment current from any one receiver is small and has little effect on the output voltages of the transmitter. Where this is done the stiffness of the receiver is small, and it is suitable for operating only a light pointer. Precautions must be taken to minimise frictional torques and errors due to mechanical unbalance of the receiver rotor.

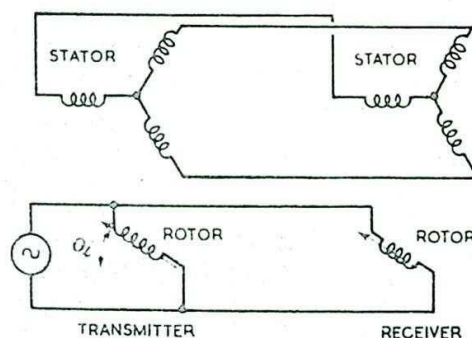


Fig. 861 - Three-circuit selsyn system.

9. Construction of Selsyn Transmitters

The principal object in the construction of a selsyn transmitter is to obtain sinusoidal variations of the mutual inductance between the primary winding and each secondary winding, as the rotor is turned. Common practice is to use three secondary windings, spaced 120° apart. Iron cores are normally used, to keep the magnetising current, at the usual power frequencies, reasonably low. Frequencies up to 1 kc/s have been used.

Usually the primary winding is put on the rotor, as this requires only two slip-rings and avoids having variable contact-resistances in the secondary circuits, the impedances of which must be

accurately balanced at all times. In addition, former-wound coils can then be used for the secondary windings. However, most of the heat losses occur in the primary, and hence some makers put the primary on the stator.

The problem of making the flux linkages between the primary and each secondary winding vary sinusoidally with the angular rotation of the rotor is the same as that of securing a sinusoidal output voltage waveform from an alternator, and the methods of construction adopted in the two cases are similar.

10. Selsyn Receivers

The receiver of a selsyn-type transmission system should resemble the transmitter in having the secondary coils so wound that the mutual inductance between each secondary winding and the primary, or polarising, winding varies sinusoidally as the rotor is turned. It is an advantage if the rotor is magnetically symmetrical.

Consider a two-circuit receiver. If the rotor is unpolarised but magnetically asymmetrical, a current flowing in either stator coil tends to align the rotor with its axis of least reluctance in the direction of the magnetic axis of the coil; i.e., the rotor tends to set in the position giving maximum inductance of the coil. If the current in the coil is I_1 (RMS value) and its inductance is L_1 , the torque on the rotor satisfies the relation

$$\tau_1 \propto \frac{I_1^2}{2} \cdot \frac{dL_1}{d\theta_0}, \quad \theta_0 \text{ being the angular position of}$$

the rotor. A similar torque $\tau_2 \propto \frac{I_2^2}{2} \cdot \frac{dL_2}{d\theta_0}$ is developed between the second coil and the rotor, so that for equilibrium,

$$\tau_1 + \tau_2 = 0.$$

This gives

$$\frac{dL_1}{d\theta_0} / \frac{dL_2}{d\theta_0} = - \left(\frac{I_2}{I_1} \right)^2.$$

This does not lead to a simple expression for θ_0 . In general the torques due to rotor asymmetry are not likely to tend to produce the same equilibrium position for the rotor as those due to mutual inductance between rotor and stator coils. Hence it is usually desirable for the rotor to be symmetrical so that $\frac{dL_1}{d\theta_0}$ and $\frac{dL_2}{d\theta_0}$ are zero. In this case the same design considerations apply to the receiver as apply to the selsyn transmitter; (Sec. 9).

11. The Magslip receiver

Simplified plan and elevation drawings of a magslip receiver are shown in Fig. 862. The I-shaped soft-iron rotor is balanced so that its equilibrium position is independent of gravity. It is polarised from a fixed cylindrical polarising winding as shown. The stator has three sets of coils arranged with their magnetic axes spaced 120° apart. The mutual inductance between any stator winding and the polarising coil should vary sinusoidally with the rotation of the rotor; i.e., the flux

entering the rotor due to any stator current should vary sinusoidally with the angular position of the rotor. This desired variation of the rotor flux can be achieved if the magnetic potential U of the

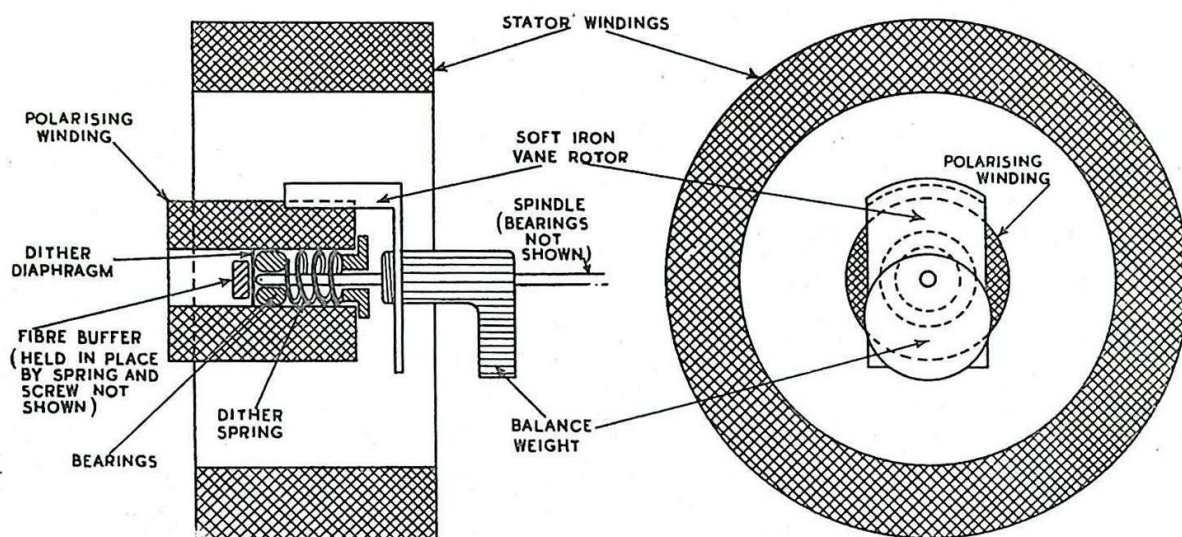


Fig. 862 - Magslip receiver.

stator surface varies sinusoidally from point to point. Using a slotted stator, this condition can be approximated to by an appropriate distribution of ampere turns. The same current flows in all the conductors of any one winding, but the number of turns in each slot is so chosen that the resulting stepped graph of magnetic potential is approximately sinusoidal; (Fig. 863). The integrating effect due to the rotor overlapping more than one slot smooths out the flux-linkages so that the effective magnetic potential is less irregular than that of Fig. 863 (b).

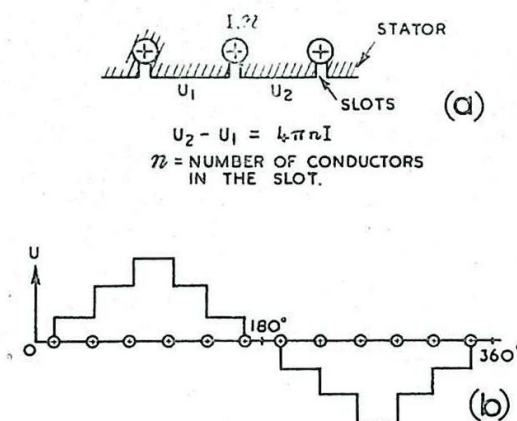


Fig. 863 - Distribution of magnetic potential for 14-slot stator.

When each of the stator windings is energised the resultant potential distribution at the stator surface is still approximately sinusoidal, as may be readily deduced by vectorial or analytical methods. There is in practice a cyclic error of magnitude approximately

6' and of period equal to the slot pitch (15° in the magalip receiver).

The effects of friction (see Sec. 15) are greatly reduced by the endwise dither of the rotor which is caused by the pulsating magnetic pull of the polarising coil.

The resonant frequency of a magalip receiver is in the neighbourhood of 2.5 c/s. The velocity lag is indicated in Fig. 864; it is about 1° at $20^\circ/\text{sec}$. The damping is low. Various methods are being investigated of increasing the damping and reducing the pronounced resonance which has deleterious effects when used with some computing devices.

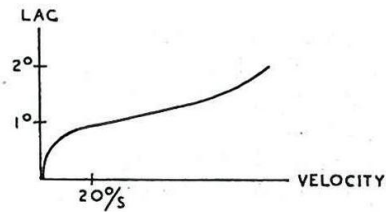


Fig. 864 - Magalip velocity lag.

12. Coincidence Indicating Systems

The object of a Coincidence Indicating System is to indicate when one shaft at the receiving end of the system is aligned with another at the transmitting end. Usually the magnitude and direction of any misalignment must be shown over a certain misalignment range.

A ring-potentiometer coincidence system is shown in Fig. 865(a). The voltmeter shows a reading whenever the angle between the two brush sets is other than 90° . The action is illustrated in Fig. 865 (b). θ denotes angular rotation from the diameter AC on either ring-potentiometer, θ_1 being the rotation of the feedpoints of voltage, $+V$ and $-V$, and θ_0 the

rotation of PQ from AC.

Neglecting the current drain from the first network due to the second, the variation of voltage with θ_0 is shown in

Fig. 865(b) for the chosen value of θ_1 . Provided $\theta_0 = \theta_1 + 90^\circ$ the voltage difference between P and Q is zero.

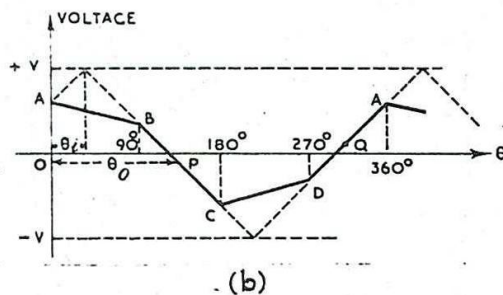
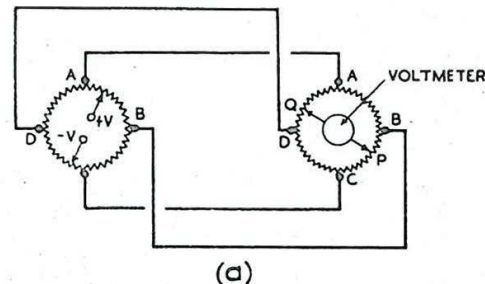


Fig. 865 - Ring-potentiometer coincidence system with curve of voltage against angular rotation.

The voltmeter may be at the same end as the power supply. It must be polarised if the sense of the misalignment is to be indicated. The system will operate with an alternating supply provided the voltmeter is polarised from an in-phase supply.

Fig. 866 shows a selsyn coincidence indicator. The flux-distribution due to rotor 1 is reproduced by stator 2. Unless rotor 2 is perpendicular to this flux, and therefore to rotor 1, a resultant EMF will be induced in rotor 2 and will be indicated by the voltmeter. The meter must be polarised, as indicated, from an in-phase supply. Although a dynamometer meter is shown, an AC-polarised moving iron meter could also be used, or a polarised DC voltmeter in conjunction with a phase discriminating rectifier; (Sec. 14).

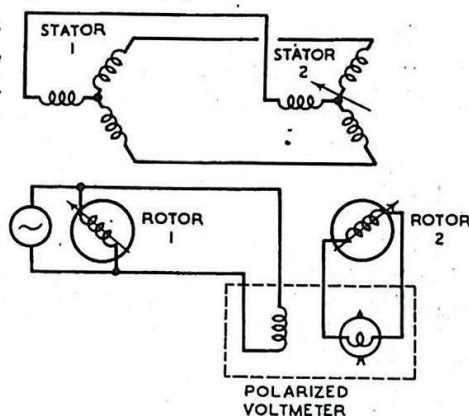


Fig. 866 - Selsyn-type coincidence system.

A selsyn system using a differential receiver is shown in Fig. 867. A differential selsyn receiver consists of a stator and a rotor, each with three symmetrical windings. The flux in the stator of the receiver is parallel to the magnetic axis of rotor 1, at an angle θ_i from the reference axis. The rotor of the receiver, carrying the rotor coils, aligns itself so that the stator flux and rotor flux axes are coincident. The rotor flux is inclined at an angle θ_o to the

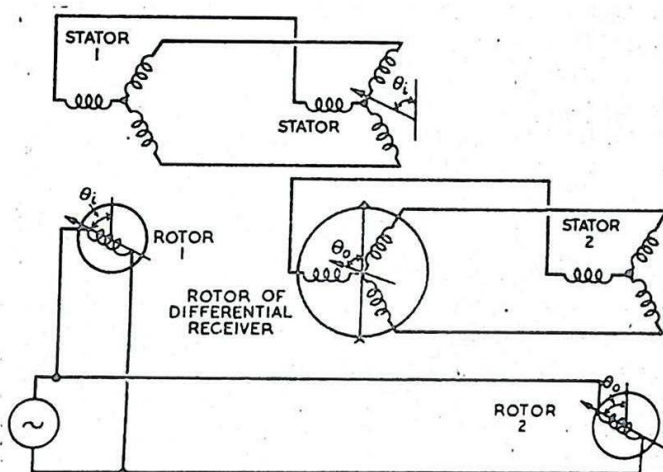


Fig. 867 - System using differential receiver.

reference axis on the rotor (indicated by the needle attached to the rotor). Hence the inclination of the rotor needle to the reference axis of stator 1 indicates the error $(\theta_i - \theta_o)$. In this system the differential receiver may be remote from either transmitter.

Coincidence indicating systems may be used to give visual indication of misalignment, or may be used, if suitably adapted, as error indicating devices in servo mechanisms.

SERVO SYSTEMS

13. General

Although attention is confined, in the subsequent analysis and description, to servos which drive a mechanical load, the principles are the same if the output is of a different kind, and the same methods of analysis may be employed.

The arrangements of the principal elements in a servo are shown schematically in Fig. 868. Sometimes the functions of two or more elements may be performed by a single co-ordinating element. It will be assumed, unless otherwise stated, that there is no inherent feedback from output to input in any individual element.

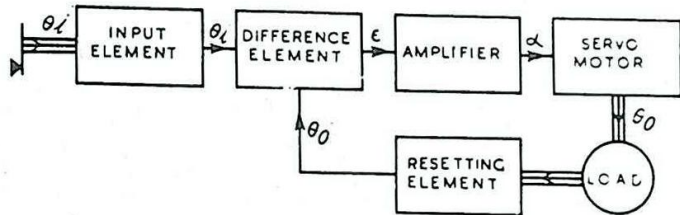


Fig. 868 - Arrangement of the principal elements of a servo (see Sec.18 for explanation of symbols).

14. Computing Devices (Converting and Calculating Elements)

In many servos it is necessary to convert operative quantities from one kind to another. For example, either the input or the error quantity may have to be converted from a mechanical movement to an electrical potential, or vice versa. It may also be required to add or subtract, multiply or divide, integrate or differentiate, various operative quantities. Any of these processes may be performed in various ways, either mechanical or electrical solutions being common.

To convert a mechanical movement to an electrical potential difference a simple linear potentiometer arrangement will suffice, as illustrated in Fig. 869. A constant voltage is applied at the input terminals A and B and the output voltage between the slider C and B is proportional to the product of the constant voltage and the slider movement. If the input voltage is not constant, but is proportional, say, to x , whilst the slider movement is proportional to y , the output is proportional to the product xy . By winding the potentiometer according to a non-linear law so that $R=f(y)$, where R is the variable resistance and y the slider movement, the output can be made proportional to $xf(y)$.

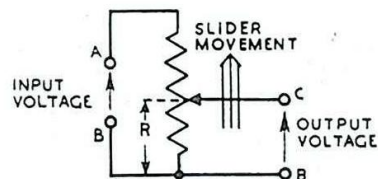


Fig. 869 - Conversion of a mechanical movement to an electric potential.

In converting from an electrical potential to a mechanical movement, a self-contained servo is usually required, since a voltmeter type of arrangement, which would otherwise suffice, does not normally provide sufficient output torque.

The fundamental arrangement is illustrated schematically in Fig. 870. The output from the resetting potentiometer is subtracted from the input in a difference amplifier. The latter controls the servo motor which drives the resetter until the input to the difference amplifier is zero. Thus the output of the potentiometer is equal to the input, and the resultant movement of the potentiometer arm is proportional to the input voltage.

Voltages may be added by using the network of Fig. 871. The result

$$V = \frac{\sum \frac{V_i}{R_i}}{\sum \frac{1}{R_i}} \quad (\text{see Chap. 1 Sec. 12})$$

becomes, if all the resistances are equal,

$$V = \frac{\sum V_i}{n} \quad \text{where } n \text{ is the number of shunt networks.}$$

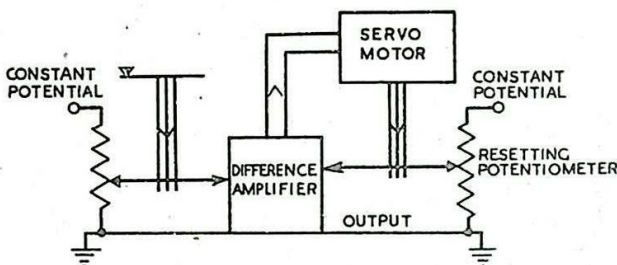


Fig. 870 - Conversion of an electric potential to a mechanical movement.

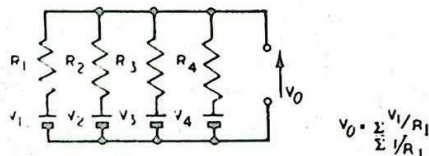


Fig. 871 - Adding network (ladder).

For direct voltages, a reversal of polarity of one of the sources changes the addition to a subtraction. Fig. 872 shows how the potentiometer input and output elements may employ this principle to form the combined resetting link and difference element of a servo. In this example the input element is a resistive potentiometer fed from +300V, the slider being rotated by the input operator. The resetting element consists of another potentiometer fed from -300V, its slider being driven by the output shaft of the servo motor. The difference voltage \mathcal{E} is taken from the junction of the two equal resistances and may be positive or negative, according as $\theta_1 \geq \theta_0$. If it is important that \mathcal{E} should accurately represent $\theta_1 - \theta_0$, the output impedance of the potentiometer should be $\ll R$. In the case illustrated this is usually unimportant, since strict proportionality is not necessary in a difference

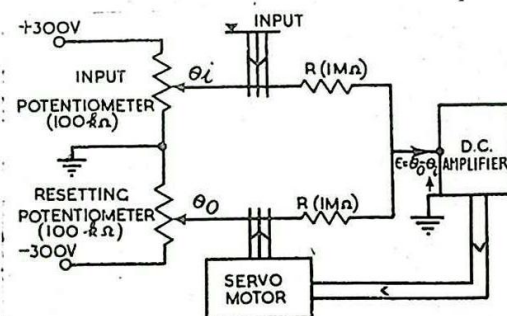


Fig. 872 - Use of potentiometers and resistance network to form combined resetting and difference elements of a servo.

element, although any non-linearity which may be introduced complicates analysis. In other cases an amplifier inserted between each potentiometer and the adding network will ensure that the output impedance is constant to a sufficient degree of approximation.

For alternating voltages the same system may be used, subtraction being achieved by a reversal of phase of one of the input signals.

Alternatively, transformers may be employed, making addition and subtraction extremely simple in theory, as indicated in Fig. 873. In practice, slight changes in phase may be difficult to avoid, and can seriously affect operation. With AC,

potentiometers may be of the inductive type, in which coils, instead of resistance-wire, wound on high permeability cores, take the place of the tapped resistance. One advantage of this arrangement is that with resistive loads the load current is in quadrature with the potentiometer current and does not appreciably affect the magnitude of the output voltage. An amplifier following the potentiometer is thus often avoidable.

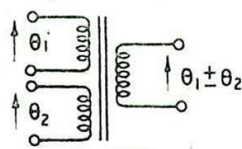


Fig. 873 - Addition or subtraction of sinusoidal voltages.

Electrical integrating and differentiating circuits are described in Chap. 2 Sec. 17. Their mechanical equivalents, in the form of ball-and-disc gear, are not often used in servos, although the "governor" type of speed control, which is a form of rate-measurer, may be employed. Integrating circuits in hydraulic servos are common, and one type consists of a cylinder with some form of valve which allows fluid to enter at a rate proportional to the size of the aperture; the output depends on the movement of a piston which is proportional to the quantity of fluid in the chamber. The piston movement is thereby made proportional to the time-integral of the valve movement.

It may be necessary in electrical servos to convert from AC to DC or vice versa. In some cases the error voltage is sinusoidal, whereas the servo motor requires a steady voltage or direct current feed, its sign determining the direction of rotation of the armature. The sign of the error depends on the phase of the alternating voltage, and it is therefore necessary to convert the sinusoidal input to a DC output by means of a phase discriminating circuit.

This may be accomplished by a synchronous rotary converter, the action of which is illustrated in Fig. 874. The commutator, which acts simply as a reversing switch, is driven at synchronous speed with reference to the frequency of the sinusoidal error voltage. At (a) is shown a fictitious reference voltage corresponding to the commutator rotation, and (b) shows the error voltage which is fed to the commutator via slip-rings. A reversal of the error voltage reverses the polarity of the brushes (d) and thus the sign of the output voltage after being filtered from the commutator.

An electronic circuit which accomplishes a similar result is illustrated in Fig. 875 (a). The standard reference voltage, with which the error voltage is either in phase or antiphase, is applied to the anodes of the two valves in pushpull, so that one is conducting when the current of the other is cut off. The error voltage is applied to the control grids of both valves, so that anode current is

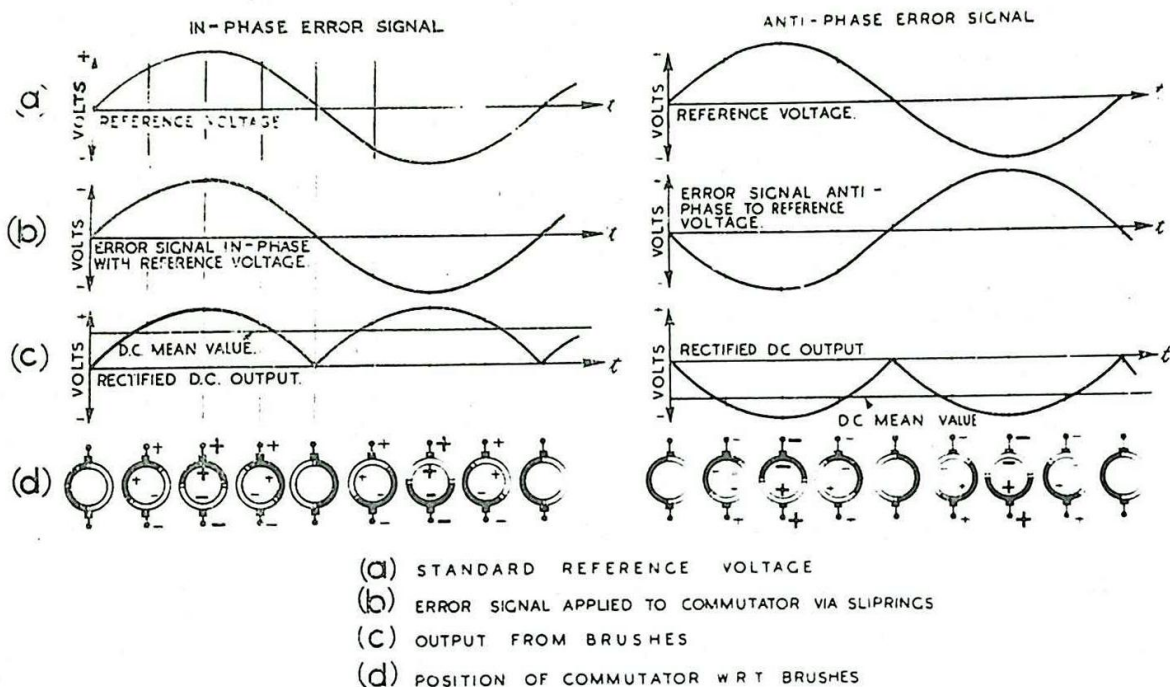


Fig. 874 - Phase discrimination by means of a synchronous rotary rectifier.

caused to flow in one valve or the other for a half cycle, as shown in Fig. 875(b). The direction of the field in the anode coil is thus reversed with the change in phase of the error voltage and, provided the linear portions of the valve characteristics are used, the magnitude of the current will be proportional to the magnitude of the error signal. In the case illustrated a direct current is caused to flow in one or other of the differential field windings of a servo motor, according to whether the error voltage is in-phase or anti-phase with the reference voltage.

The reciprocal problem of converting a DC input of a certain polarity into an output of corresponding amplitude and phase may be solved by means of saturable reactors as well as by electronic means. Fig. 876 (a) shows an iron-cored reactor fed with both direct and alternating voltages. The direct current partially saturates the core, so that the impedance presented to the AC input terminals is reduced as the direct current is increased.

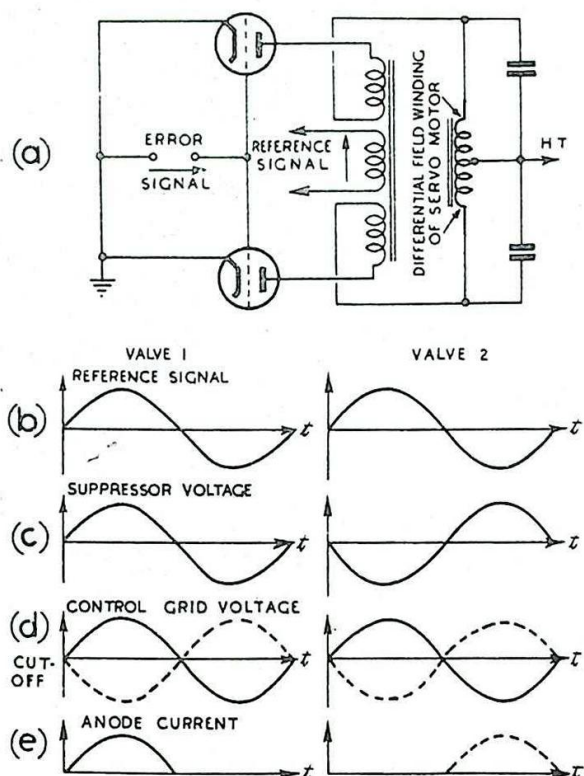


Fig. 875 - Operation of electronic AC-DC error converter (Phase-discriminating rectifier).

Four such reactors are used in the network shown in Fig. 876(b). DC polarising windings, one on each reactor, are fed from the same source. Control windings, fed from the DC error voltage source, are so connected that the DC fields in A and C are increased by the error voltage when those in B and D are decreased, and vice versa. When the error voltage is zero and the system is properly balanced, the impedances presented by the four reactors at the individual AC input terminals are the same. When the error voltage is not zero the DC fields in two opposite arms of the bridge are increased, reducing the AC impedance, and those in the other pair of arms are reduced, increasing the AC impedance. Thus a net voltage is produced at the output terminals equal to

$$\frac{V_i (z_A - z_B)}{z_A + z_B}$$

$$z_A + z_B$$

and is thus in-phase or anti-phase with V_i according to the polarity of the DC error signal. Within limits the relation is linear.

15. Characteristics of the Load

Where the servo is required to provide mechanical movement the dynamic characteristics of the load, motor armature and gearing are of primary importance in their effect on the behaviour of the system as a whole.

The principal effects to be considered are inertia and friction. The former is the chief limiting factor with respect to acceleration or retardation, and the latter dissipates energy and limits the speed

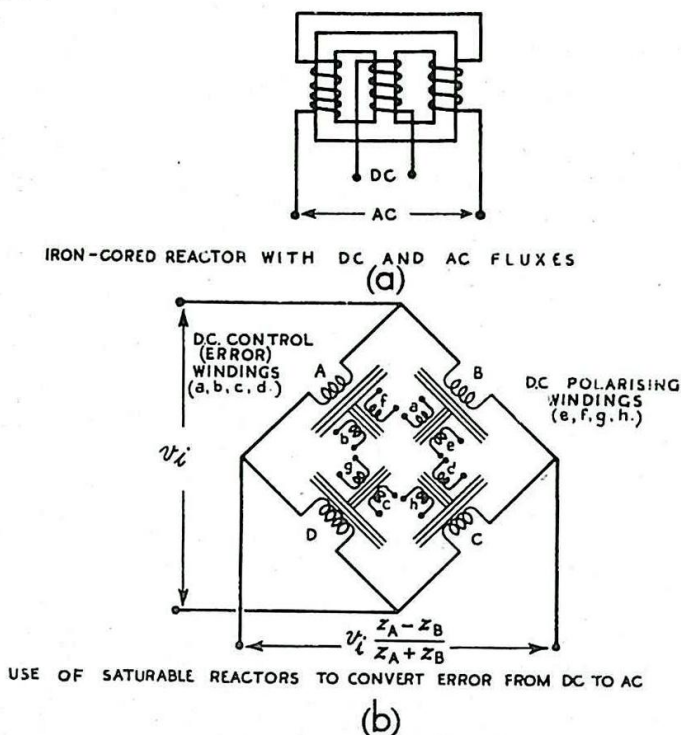


Fig. 876 - Iron-cored reactor with DC and AC fluxes, and use of saturable reactor to convert error from DC to AC.

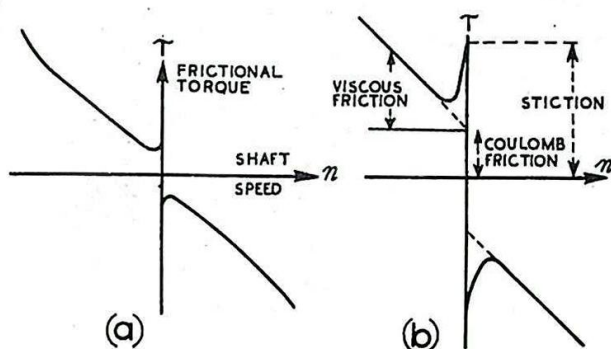


Fig. 877 - Variation of frictional torque with shaft speed.

of rotation. Fig. 877 shows the variation of frictional resistance or torque with shaft speed. There are three chief components of the friction force, stiction (or static friction), coulomb friction and viscous friction. These are indicated at Fig. 877(b). In most analyses of servos the first two types are ignored, since they are non-linear and difficult to allow for. The co-efficient of viscous friction is sometimes called Viscosity. In practice, stiction is particularly important, its presence leading to jerky motion (since the stiction torque must be overcome before motion can ensue, and this involves an error in the input to the servo motor before sufficient torque can be built up).

Various methods are used to combat stiction, one of the commonest being to dither the output shaft either by using an auxiliary motor, or by introducing a deliberate dither error into the servo mechanism. This keeps the shaft moving at a rapid rate about the equilibrium position, and provided the dither frequency is high enough (of the order of 10-1000 c/s, depending on the natural frequency of the system) the effective resistance to be overcome to cause initial motion is reduced from its stiction value almost to zero. The dither should be in a direction perpendicular to the plane of motion: e.g. a magstrip receiver pointer (see Sec. 11) is caused to dither in the direction of the axis of rotation. A rod which moves in a longitudinal bearing might be given a rotational dither motion. Another method of counteracting stiction is that of variable duration impulsing. This is similar to the principle of the piledriver. Stiction may be sufficient to prevent motion if only a steady relatively small force is applied. If a much greater force is applied for short periods, the duration of each impulse being variable, no matter how small the mean force applied the peak force may be maintained sufficiently large to make stiction unimportant. The use of variable duration impulsing is illustrated in Fig. 878. The momentum imparted to the load is proportional to the excess of the area below the impulse curve over the corresponding area below the friction curve. Once motion has begun it is the mean frictional torque of Fig. 878 that matters, not the stiction value.

16. Servo Motors

Many forms of DC or AC machines may be used in servo systems, besides other types of motors in common use, and a full description of them is impossible in this work. A few of these will be mentioned with special reference to their servo applications.

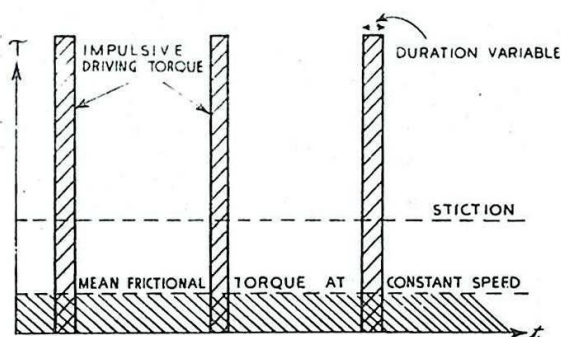


Fig. 878 - Use of variable duration impulsing to combat stiction.

In general the motor is required to remain stationary (for a displacement-controlling-displacement servo) when some controlling current or voltage is zero, and to accelerate as this error quantity is increased, in a direction depending on the polarity (DC) or phase (AC) of the error. The motor may also possess additional features, such as automatic damping or braking devices, other than frictional resistance at the rotor bearings.

The simplest of servo motors are of the On-Off type, there being no variation of motor speed or torque with the amplitude of the

error voltage. Such motors are prone to excessive Hunting; i.e., the output shaft over-runs the equilibrium or zero-error position, the drive to the motor is reversed, and the output shaft again over-runs in the opposite sense, the output thus being of an irregular, oscillatory nature. Alternatively there is likely to exist a substantial Dead Zone; i.e., the error must exceed a certain value, either positive or negative, before the motor is switched on. For accurate displacement-control servos such on-off motors are seldom desirable, and a smoothly variable type of control is necessary. In the analysis of such systems, it is generally assumed that either the output torque or the shaft speed of the motor is approximately proportional to the magnitude of the input voltage or current. Whilst it is true that many types of motor do behave in one of these ways, in general the relation between output and input is much more complicated than this. It will, however, be shown how a DC motor, by means of a separate feedback circuit, can be made to conform very approximately to the Velocity Control requirement that speed of the output shaft is proportional to the control voltage, thereby simplifying design. Where it is the output torque, rather than the output speed, which is dependent on the magnitude of the error voltage or current, the term Torque Control is applied.

DC Motors

There are two methods of controlling a DC motor with a separately excited field winding. The control voltage may be applied either to the armature or to the field winding. The advantage of the former is that when there is no output required i.e., when the motor is at rest, no power is dissipated in the armature circuit since no current flows there. The disadvantage lies in providing a supply for the armature current with a sufficiently low output impedance. For high-powered motors gas-filled control valves of the thyatron type are frequently used, and these have other drawbacks (see Chap. 6) apart from their comparative fragility. Other types of current amplifiers of a non-electronic nature are sometimes used. One of these, used in the Ward-Leonard system, is described below.

One method of controlling a DC motor by means of the field winding employs a reactor of large inductance in series with a bridge rectifier circuit feeding the armature winding. While the armature is stationary no armature reaction EMF is generated, and the AC armature supply circuit is almost purely reactive. When an error voltage develops a field, the armature accelerates in the direction corresponding to the polarity of the error, and power is drawn from the armature supply due to the back EMF, which makes the armature circuit partly resistive.

In general, series or shunt field arrangements are not suitable for use in servo motors, since a reversal of the armature voltage does not result in a reversal of the output torque. A differential-field series motor may be used, with opposing field-windings switched either mechanically or electronically by the error signal.

Two-Phase Induction Motor

Many types of this motor (using cup-shaped metal rotors, or squirrel-cage or wound armatures) are in common use. The method of control normally used is to apply a constant alternating voltage to one stator field-winding and the error voltage, in quadrature with the constant voltage, to the other winding. The motor will then accelerate in one direction or the other according to whether the error voltage leads or lags in quadrature with the constant voltage. Also, since the torque is proportional to the currents induced in the armature, which in turn depend on the magnitude of the rotating flux, the larger the error voltage the greater is the output torque. A special point in design is that when the error voltage is reduced to zero the motor will hunt if allowed to

run as a single phase motor. This may be prevented by a suitable choice of reactance/resistance ratio for the armature winding. When properly designed, the armature acts as an eddy current brake when the error voltage is zero, and this feature makes the two-phase motor particularly adaptable as a servo motor.

Inductance motors used in this way are not economical except at low powers, and it is in small servos that they are usually to be found.

Ward-Leonard System

This high-powered type of control gear consists of a three-element chain of prime-mover, DC generator and variable speed motor. The arrangement is illustrated in Fig. 879. The first two elements act as a current amplifier, since the power supplied to the variable speed motor depends on the current in the exciting field-winding. Various modifications exist which affect the behaviour of the system, particularly in regard to the field winding of the generator.

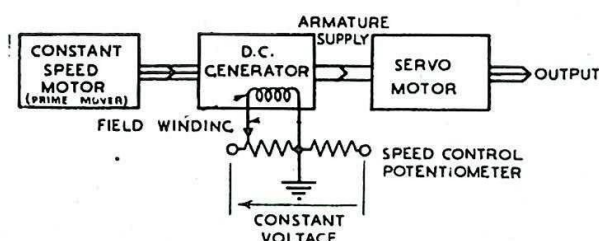


Fig. 879 - Ward-Leonard system.

The metadyne is a DC generator used as in the Ward-Leonard system for driving a DC servo motor. The characteristics of the whole system depend largely on the amplification of the generator and its rapidity of response to changes in the controlling field-current. By suitable design the metadyne can be made to produce very considerable power amplification, of the order 30,000 : 1, but the same elements of design which provide high amplification limit the rapidity of response. When the metadyne is designed to give maximum power amplification it is known as an Amplidyne Generator.

One of the characteristics of a Ward Leonard system using a metadyne is that a braking torque may be automatically developed when the speed of the output shaft of the motor load exceeds the speed demanded by the magnitude of the error. This is a useful anti-hunt feature, since the brake is applied in this case before the output overshoots, i.e. before the error reaches zero.

Speed-Control System

Fig. 880 illustrates a method whereby a DC motor can be made to conform approximately to the speed-control type; i.e. its output shaft speed is proportional to the magnitude of the error voltage irrespective of the output torque, within the limits of the linearity of the motor characteristics.

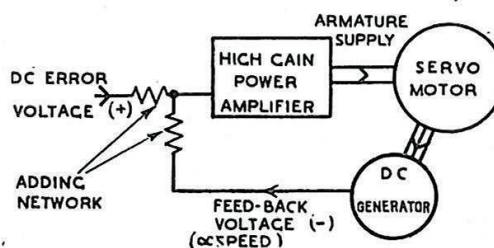


Fig. 880 - Speed control system.

A small DC generator is attached, or geared, to the output shaft. The output voltage from this generator is

proportional to shaft speed within a close degree of approximation. This voltage is subtracted from the error voltage in the amplifier input circuit and the difference is used to control the DC motor. The change in torque thereby produced at the output shaft is very much more rapid than the corresponding change in error voltage, and, provided the motor does not overload, the difference between the error and feedback voltages is kept very small. Hence the error, as well as the feedback, is very nearly proportional to the speed of the output shaft.

17. Performance of Servo Systems

No standard technique has yet been adopted for measuring or estimating the complete performance of a servo system, but various tests which can be applied will be dealt with in this section. Possibly the simplest complete test which can be devised is the frequency response. A simple harmonic input is applied, and the amplitude and phase of the output are determined. The phase difference and relative amplitude for different frequencies may then be plotted on a harmonic response diagram as shown in Fig. 881. Because some elements of the preceding or succeeding stage of radar control systems may exhibit pronounced resonant properties it is frequently important not only to avoid resonances, but to introduce deliberately a small response in the servo at the resonant frequencies of the external circuits. It may be sufficient if the response to frequencies higher than one or two c/s falls off very rapidly with rising frequency, but this also tends to make for sluggish response to sudden changes of input. Where the input quantity to a servo is a received radar signal, which is prone to certain types of unwanted fluctuations, a special frequency response may be required.

A second test which may be applied to a servo is its response to a sudden change of input quantity; (Unit Function). If the input quantity is altered suddenly by unit

amount from the equilibrium position the manner in which the output approaches its new value is called the Unit Function Response. Complete information as to the performance of the servo to any given input can be derived from this test; (provided the servo is a linear system). In particular the rapidity of pull-in and whether it is oscillatory or non-oscillatory, and of what degree, are of considerable importance in most radar applications.

Other tests aim at subjecting the servo to such input conditions as are more directly related to the actual conditions likely to be met with in use. There is usually a maximum input velocity which can be expected and at this velocity of input there must not be introduced a velocity lag greater than the permissible error. Similarly, the acceleration and higher order lags must not exceed stipulated values, and tests may be applied to confirm this. There is liable to be a

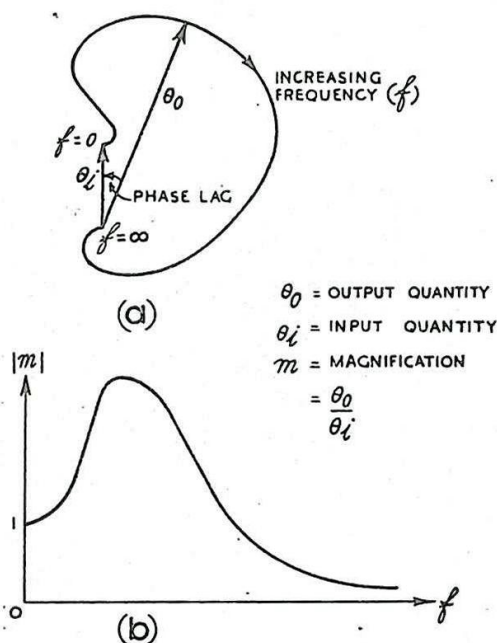


Fig. 881 - Typical harmonic response diagrams.

standstill error if the system is insufficiently stiff. If the system is too resilient, i.e., not stiff enough, extraneous torques due to windage, stiction etc. may introduce excessive errors.

18. Definitions and Assumptions (Applying to a displacement-displacement servo).

The input, output and error quantities are denoted by Θ_i , Θ_o , and E respectively, so that $E = \Theta_i - \Theta_o$. These are converted by gearing into proportionate quantities of the same kind - i.e., an angular rotation - so that the new quantities, θ_i , θ_o and ϵ are connected with the old by the relations

$$\frac{\theta_i}{\Theta_i} = \frac{\theta_o}{\Theta_o} = \frac{\epsilon}{E} = c \text{ where } c \text{ is the gear ratio.}$$

(See Fig. 882).

The input and resetting elements convert Θ_i and Θ_o into voltages or currents which are combined in the difference element, the difference being amplified and applied, either directly or through some modifying circuit, to the servo motor. The input to the motor is denoted by $\alpha = f(\theta_i, \theta_o)$ and it will be assumed that the output torque of the motor is T_o where $T_o = k_o \alpha$

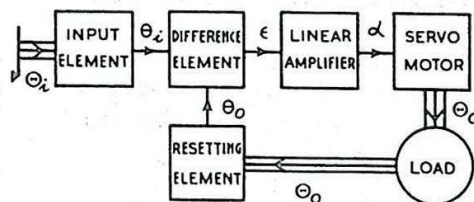


Fig. 882 - Simple error control.

The gear ratio c is often sufficiently high to ensure that the mechanical properties of the load may be neglected in comparison with those of the gearing and motor shaft. If this is not so the moments of inertia and coefficients of friction for the load may be referred to the output shaft of the motor according to the gear ratio and added to the mechanical properties of the motor.

The resultant coefficient of viscous friction will be denoted by k and the moment of inertia by J . Coulomb friction and stiction, time-lags and backlash will be neglected. For some purposes we shall assume that the output shaft is subjected to an extraneous torque T_x , due, for example to windage or stiction.

19. Simple Error Control

In this case, illustrated by the circuit of Fig. 882, the input current or voltage applied to the motor is proportional to the error E .

i.e. $\alpha = \gamma \epsilon$, where γ is a constant, and

$$T_o = k_o \alpha = k_o \gamma \epsilon = k_o \gamma c E$$

k_o is a constant for a given motor, but γ depends on the amplification provided between the difference element and the motor.

The net torque produced at the rotor shaft is therefore

$k_o \gamma \mathcal{E} + T_x - \frac{k d\theta_o}{dt}$, and this is equal to the rate of change of angular momentum, $J \frac{d^2\theta_o}{dt^2}$

$$\text{i.e., } k_o \gamma \mathcal{E} + T_x - k \frac{d\theta_o}{dt} = J \frac{d^2\theta_o}{dt^2}$$

$$\text{i.e., } k_o \gamma (\theta_i - \theta_o) + T_x = J \frac{d^2\theta_o}{dt^2} + k \frac{d\theta_o}{dt}$$

$$\text{Hence } J \frac{d^2\theta_o}{dt^2} + k \frac{d\theta_o}{dt} + k_o \gamma \theta_o = k_o \gamma \theta_i + T_x \dots\dots\dots (1)$$

Putting $\theta_o = \theta_i - \mathcal{E}$, and rearranging the terms, we obtain

$$J \frac{d^2\mathcal{E}}{dt^2} + k \frac{d\mathcal{E}}{dt} + k_o \gamma \mathcal{E} = J \frac{d^2\theta_i}{dt^2} + k \frac{d\theta_i}{dt} - T_x \dots\dots\dots (2)$$

These two linear differential equations form the basis of the analysis of the system.

Steady State Errors

We shall assume for the moment that a steady state can exist in each of the forms which we are about to consider. This assumption presupposes the stability of the system, which will be discussed later. By steady state we mean here a condition in which the input obeys a simple law, such as a constant velocity, or constant acceleration, while the output follows a related law with an angular lag corresponding to the magnitude of the input. We shall consider the three conditions, $\theta_i = \text{constant}$, $\frac{d\theta_i}{dt} = \text{constant}$, and $\frac{d^2\theta_i}{dt^2} = \text{constant}$.

Stiffness

If θ_i and θ_o are both constant, equation (2) gives

$$\mathcal{E} = \frac{-T_x}{k_o \gamma} \quad \text{or } k_o \gamma = - \frac{T_x}{\mathcal{E}}$$

$k_o \gamma$ is called the Stiffness Coefficient of the servo, since it determines the magnitude of the extraneous torque which must be applied to the motor shaft to produce a given error \mathcal{E} . The corresponding coefficient for the load, as distinct from the motor shaft, is $c^2 k_o \gamma$, since the torque required is increased, and the angular error decreased, in the ratio $c:1$.

Velocity Lag

If the input velocity is constant, in the steady state

$$\frac{d\theta_i}{dt} = \frac{d\theta_o}{dt} = A, \quad \text{say;}$$

then $k_o \gamma \varepsilon = kA - \dot{T}_x$, from (2);

$$\text{i.e. } \varepsilon = \frac{kA}{k_o \gamma} - \frac{\dot{T}_x}{k_o \gamma}.$$

$\frac{k}{k_o \gamma}$ is called the velocity lag coefficient; (i.e., it is the lag per unit constant velocity).

Acceleration Lag

$$\text{If } \frac{d^2 \theta_i}{dt^2} = B$$

$$\frac{d\theta_i}{dt} = Bt + A = \Omega, \text{ say.}$$

It can be shown from equation (2) that

$$\varepsilon = \left(\frac{J}{k_o \gamma} - \frac{k^2}{k_o^2 \gamma^2} \right) B + \frac{k}{k_o \gamma} \Omega - \frac{\dot{T}_x}{k_o \gamma} \text{ in the steady}$$

state.

$$\frac{J}{k_o \gamma} - \frac{k^2}{k_o^2 \gamma^2} \text{ is called the acceleration lag coefficient.}$$

(It is immaterial whether velocity or acceleration lag coefficients are referred to the motor output shaft or to the load itself, since both error and velocity or acceleration are similarly changed by the gear ratio. Also the same figure is obtained whether the lag is quoted in radians per radian-per-second or in degrees per degree-per-second).

Using the notation of a subsequent paragraph, we may write the acceleration lag coefficient as $\frac{J}{\gamma k_o} (1 - 4\zeta^2)$, where $\zeta = \frac{k}{2\sqrt{k_o \gamma J}}$.

In a manner similar to the above, lag coefficients of higher orders may be obtained, if required. In practical applications however, it is usually the lower order lags which predominate, and if these are kept within the required minimum the total errors remain sufficiently small.

Response to Unit Function Input ($\dot{T}_x = 0$)

If θ_i follows the variation shown in Fig. 883(a) the response of θ_o may be oscillatory or non-oscillatory as shown at (b) according to the ratios of the coefficients of the characteristic equation obtained by writing D for $\frac{d}{dt}$ in equation (1) and equating to zero the coefficient of θ_o .

$$\text{This gives } JD^2 + kD + k_o \gamma = 0.$$

This may be written $D^2 + 2\zeta\omega_n D - \omega_n^2 = 0$,

where $\omega_n = \sqrt{\frac{k_o \gamma}{J}}$, and is called the undamped natural angular frequency;

$f_n = \frac{\omega_n}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{k_o \gamma}{J}}$ is the undamped natural frequency;

$\zeta = \frac{k}{2J\omega_n} = \frac{k}{2\sqrt{k_o J \gamma}}$ and is called the Damping Ratio.

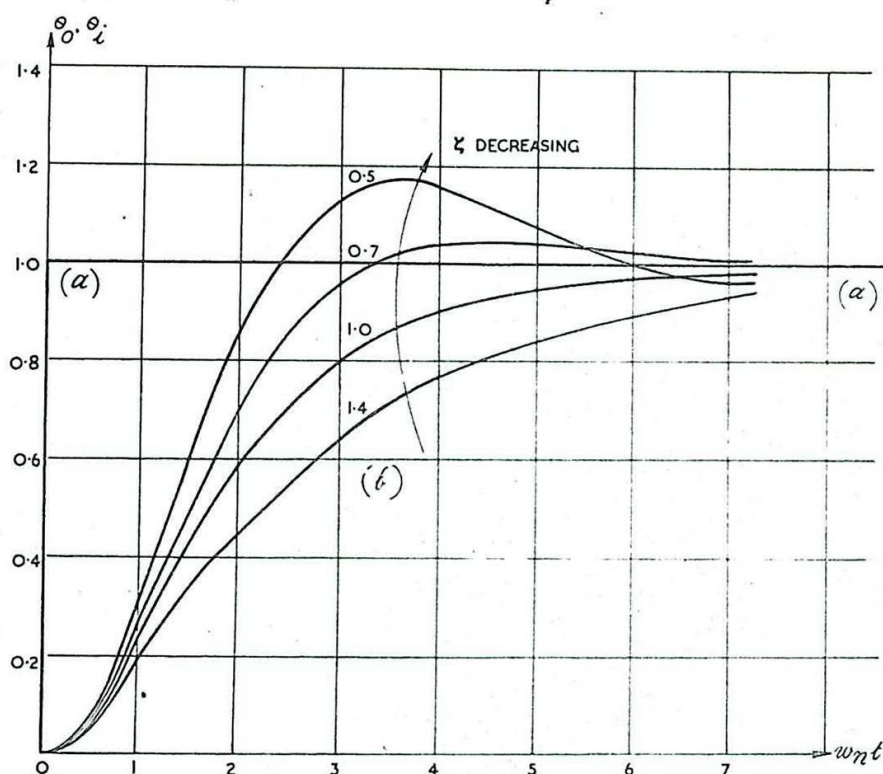


Fig. 883 - Response of simple error control servo to unit-function input.

Alternatively, $\zeta\omega_n$ may be used as a parameter, called the Damping Factor. The system is undamped if $\zeta = 0$, so that continuous oscillations are present. If $\zeta < 0$ the system is unstable (impossible in this case since neither k nor J can be negative). If $\zeta = 1$ the response is critically damped, whereas for $\zeta > 1$ it is non-oscillatory. (Compare Chap. 2 Sec.10).

Normally it is important that a markedly oscillatory response should be avoided, and in practice a value for ζ of about 0.6 to 0.8 is common. This gives a slightly oscillatory response that pulls in more rapidly than when $\zeta = 1$.

For a given value of ζ , the rapidity of pull-in depends on ω_n . The higher the natural frequency the more rapid is the response.

On the other hand there may be good reasons for keeping ω_n low, particularly because, as already stated, of undesirable resonances in other parts of the control system.

• Harmonic Response ($T_x = 0$)

If θ_i is a simple harmonic variation of frequency $f = \frac{\omega}{2\pi}$, we may obtain the response of θ_o from equation (1) using the relations

$$\frac{d\theta_i}{dt} = j\omega\theta_i \text{ and } \frac{d^2\theta_i}{dt^2} = -\omega^2\theta_i.$$

$$\begin{aligned} \text{Hence } m = \frac{\theta_o}{\theta_i} &= \frac{k_o \gamma}{k_o \gamma + jk\omega - J\omega^2} = \frac{1}{1 - \frac{J}{k_o \gamma} \omega^2 + j \frac{k}{k_o \gamma} \omega} \\ &= \frac{1}{1 - \frac{\omega^2}{\omega_n^2} + 2j\zeta \frac{\omega}{\omega_n}} \end{aligned}$$

The amplitude of the response is given by

$$|m| = \frac{1}{\sqrt{\left\{1 - \left(\frac{\omega}{\omega_n}\right)^2\right\}^2 + 4\zeta^2 \left(\frac{\omega}{\omega_n}\right)^2}} \dots\dots\dots (3)$$

Curves are plotted in Fig. 884 for different values of ζ , showing the variation of $|m|$ with x , where $x = \frac{\omega}{\omega_n}$.

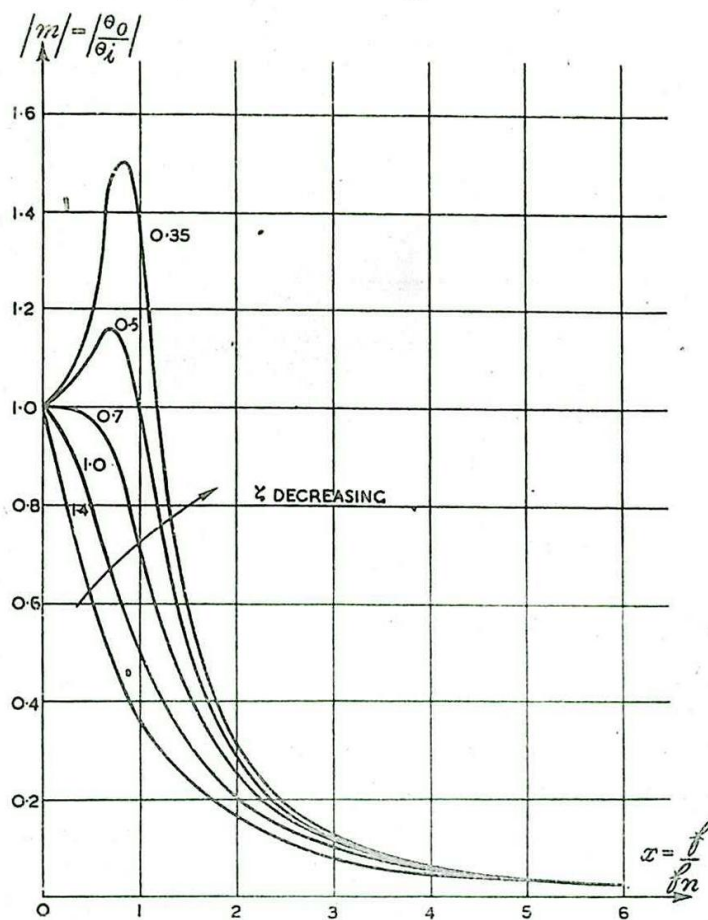


Fig. 884. - Harmonic response of simple error control servo.

20. Limitations of Simple Error Control

In this form of control system there are not normally sufficient variables to ensure that all the requirements are fulfilled. For a given motor and load k , k_0 and J are fixed (apart from minor changes which depend on the gearing) so that c and γ are the only variable parameters. It is normal to choose c so that the motor can handle the maximum speed of rotation (slewing speed) without overloading during periods of high acceleration. The other requirements all involve γ and are frequently mutually conflicting. How this works out in practice is described in the following example, taken from a radar servo providing auto-follow in bearing.

The servo motor is a $\frac{1}{2}$ HP motor with a maximum speed of

2500 r.p.m.

The maximum speed required is $10^\circ/\text{sec.}$ at the load.

Hence c is made equal to

$$\frac{2\pi \cdot 2500}{60} \div \frac{10.2\pi}{360} = 1500.$$

At this speed the output torque is given by

$$T_o = \frac{1}{2} \cdot 550 \div \frac{2500 \cdot 2\pi}{60} = \frac{33}{10\pi} \text{ lb.ft.}$$

Hence k , the coefficient of viscous friction must not be greater than

$$\frac{33}{10\pi} \div \frac{2500 \cdot 2\pi}{60} = \frac{99}{2500\pi^2} \\ \doteq 0.004 \text{ lbs.ft/rad/sec.}$$

The inertia of the rotor is 0.003 slug ft. (1 slug \doteq 32 lbs. mass).

The maximum permissible error is $\frac{1}{3}$ of a degree at the maximum speed, so that the velocity lag coefficient must not be greater than

$$\frac{1}{3} \div 10 \text{ }^\circ/\text{sec.}$$

Hence, using the formula of Sec. 19, we obtain

$$\gamma k_o > 30k.$$

For optimum damping take $\xi = 0.6$, so that

$$\frac{k}{2} = 0.6 \sqrt{k_o \gamma J}.$$

$$\text{Hence } \gamma k_o = \frac{k^2}{1.44 J}.$$

Combining these last two results, we have

$$\frac{k^2}{1.44J} > 30k$$

$$\text{i.e. } k > 30 \cdot 1.44J.$$

Putting $J = 0.003$, this becomes

$$k > 0.1296.$$

This conclusion contradicts the earlier one that k must be less than 0.004.

The damping requirement could not be satisfied unless the $\frac{1}{2}$ HP motor were replaced by a much larger one, about 14 HP being required.

Further limitations arise since the velocity lag requirements conflict with the desirability of keeping f_n small.

$$\text{Since } \gamma k_o > 30k, \text{ it follows that } f_n > \frac{1}{2\pi} \sqrt{\frac{30k}{J}}$$

and taking k as 0.004 this makes $f_n > 0.9$. (Taking k as 0.1296 makes matters far worse).

21. The Use of Derivative and Integrated Error Control

To overcome the limitations of simple error control additional terms proportional to various derivatives or time-integrated functions of the error or of the input and output quantities are included in the quantity controlling the motor. These functions may be obtained before or after the difference element; for instance, instead of θ_i being fed to the difference element directly, it may pass through various devices so that the difference element input is

$$\theta_i + A \frac{d\theta_i}{dt} + B \int \theta_i dt,$$

or some similar function. We shall use the notation $f(D)$ to indicate these additional terms, where $D = \frac{d}{dt}$; $f(D)$ may include both differential and integral forms.

The inputs to the difference element then become $\theta_i + f_i(D)\theta_i$ and $\theta_o + f_o(D)\theta_o$, so that $T_o \propto \{\theta_i + f_i(D)\theta_i - \theta_o - f_o(D)\theta_o\}$.

Fig. 885 illustrates the method.

(The output from the difference element may also be subjected to further differentiation or integration, so that

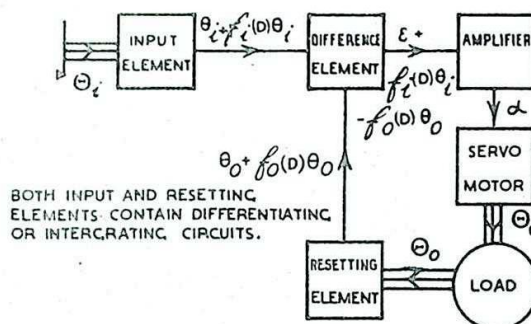


Fig. 885 - Use of derivative or integrated error control.

$$T_o \propto \{1 + f_c(D)\} \{ \theta_i + f_i(D)\theta_i - \theta_o - f_o(D)\theta_o \}$$

but for simplicity we have assumed that $f_c(D) \equiv 0$.

The differential equation of the system then becomes

$$k_o \gamma \{ \theta_i - \theta_o + f_i(D)\theta_i - f_o(D)\theta_o \} + T_x = KD\theta_o + JD^2\theta_o \equiv f_l(D)\theta_o$$

if we apply the same notation to the load function.

Hence we obtain,

$$\{k_o \gamma (1 + f_o(D)) + f_l(D)\} \theta_o = k_o \gamma \{1 + f_i(D)\} \theta_i + T_x \dots \dots \dots (4)$$

and

$$\{k_o \gamma (1 + f_o(D)) + f_l(D)\} \epsilon = k_o \gamma \{f_o(D) - f_i(D) + f_l(D)\} \theta_i - T_x \dots (5)$$

By a suitable choice of $f_0(D)$ the characteristic equation, obtained by equating to zero the coefficient of θ_0 in equation (4), may be adjusted in any desired manner without affecting the mechanical constants k and J . When this has been done the function $f_1(D)$ may be chosen to bring velocity and acceleration lags down to the required minima.

It may be useful to regard the coefficient of D in $k_0 \gamma f_0(D)$ as constituting "artificial viscosity", since it becomes added to k in the coefficient of θ_0 . This coefficient of D may be either positive or negative, so that the affect of "Negative viscosity" can be obtained. Similarly the coefficient of D^2 in $k_0 \gamma f_0(D)$ may be regarded as "artificial inertia". By the introduction of these and similar quantities the various requirements which were found to be mutually incompatible in the single system of Sec. 19 may frequently be satisfied.

There are practical reasons for avoiding the use of differentiating circuits in the input to the servo. Particularly in radar auto-follow systems the input signals are liable to be jerky, so that differentiation accentuates the irregularities. Either differentiation of the output or integration is generally to be preferred. The latter inevitably involves a time lag but in practice this can often be made negligible.

In the case of Integrated Error Control $f_0(D) \equiv f_1(D) = \frac{a}{D}$, say, where a is a constant.

We obtain from (5)

$$\left\{ k_0 \gamma \left(1 + \frac{a}{D} \right) + f_1(D) \right\} \mathcal{E} = f_1(D) \theta_1 - T_x.$$

This may be written

$$\begin{aligned} \left\{ a + k_0 \gamma D + D f_1(D) \right\} \mathcal{E} &= D f_1(D) \theta - D T_x \\ &= J \frac{d^3 \theta_1}{dt^3} + k \frac{d^2 \theta_1}{dt^2} - \frac{dT_x}{dt}. \end{aligned}$$

If θ_1 and T_x are both constant and the servo is stable, \mathcal{E} is zero in the steady state. This also follows when $\frac{d}{dt} (\theta_1)$ is constant; i.e., the velocity lag is reduced to zero. With this form of control no matter how small the error, the output torque will eventually build up (within the limits of the motor's output power) until the extraneous torque is exceeded and the error reduced.

These effects can be produced by suitable arrangement of electrical integrating or differentiating circuits such as those described in Chap. 2 Sec. 17.

Although such techniques as these allow for great flexibility of servo design, they operate under the assumption that the amplifiers, motors etc., employed in the servo remain linear under the conditions imposed by the choice of circuit components. This assumption is justified provided output torques and speeds remain well below the maximum for the motor, and provided amplifiers are not overloaded. In practice, these assumptions frequently do not apply, particularly when the system is used for slewing (turning at high speed).

22. Speed Control (First Order Servo)

If in equation (3) of Sec. 21 we substitute $f_o(D) \equiv aD$ and $f_1(D) \equiv 0$, we have

$$k_o \gamma \theta_o + a k_o \gamma \frac{d\theta_o}{dt} + k \frac{d\theta_o}{dt} + J \frac{d^2\theta_o}{dt^2} = k_o \gamma \theta_i + T_x.$$

Ignoring T_x and choosing a sufficiently large so that

$$a k_o \gamma \frac{d\theta_o}{dt} \gg \frac{k d\theta_o}{dt} + J \frac{d^2\theta_o}{dt^2}, \text{ we have}$$

$$\theta_o + \frac{a d\theta_o}{dt} \approx \theta_i \dots\dots\dots (6)$$

Substituting $\theta_o = \theta_i - \epsilon$ we have

$$\epsilon + a \frac{d\epsilon}{dt} \approx a \frac{d\theta_i}{dt} \dots\dots\dots (7)$$

This approximation will be justified in the majority of cases, exceptions being for inputs of the unit-function type, which are normally met in initial conditions only.

For such a servo the velocity lag coefficient is a , obtained by putting $\frac{d\epsilon}{dt} = 0$ in equation (7).

The harmonic response is obtained by putting

$$\frac{d\theta_o}{dt} = j\omega \theta_o \text{ so that } \theta_o = \frac{\theta_i}{1 + aj\omega}.$$

$$\text{Then } |m| = \left| \frac{\theta_o}{\theta_i} \right| = \frac{1}{\sqrt{1 + a^2\omega^2}} \dots\dots\dots (8)$$

It is clear that a rapid falling off in response with increasing frequency is obtained by making a large, i.e. at the expense of velocity lag.

The above equations are the same as those obtained if simple error control is employed with a motor whose speed, independent of output torque, is proportional to the error;

for in this case,

$$\frac{d\theta_o}{dt} \propto (\theta_i - \theta_o) = \frac{1}{a} (\theta_i - \theta_o), \text{ say,}$$

giving the same result as that of equation (6).

23. Second Order Servo with Zero Velocity Lag

If f_1 and f_o in equation (4) are suitably chosen the

equation (if T_x is ignored) reduces to :-

$$\frac{d^2 \epsilon}{dt^2} + 2\zeta \omega_n \frac{d\epsilon}{dt} + \omega_n^2 \epsilon = \frac{d^2 \theta_i}{dt^2} \dots\dots\dots (9)$$

where ζ and ω_n may be chosen without restriction. Comparing this equation with (2), we see that the term in $\frac{d\theta_i}{dt}$ has been eliminated; this reduces the velocity lag to zero. Other modifications to the response of the servo due to the disappearance of this term may be seen by comparing the response to unit-function input (Fig. 886) and the harmonic response (Fig. 887) with those of the simple error control servo (Fig. 883 and 884). The pull-in is much more rapid and for damping greater than or equal to critical ($\zeta \geq 1$) there is a

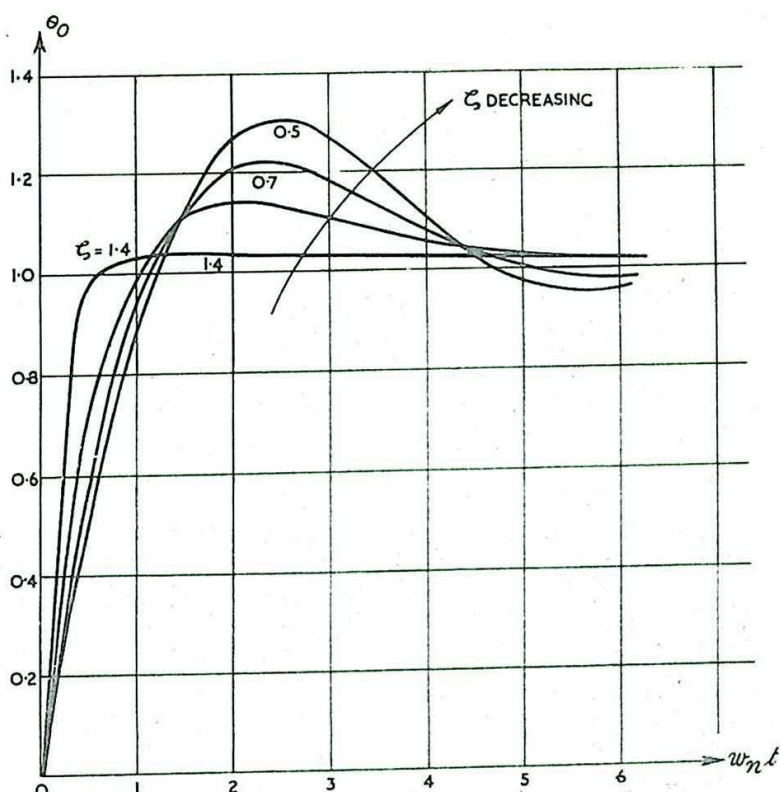


Fig. 886 - Response of second order servo (zero velocity lag) to unit-function input.

single overshoot instead of a gradual rise as in the former case. However, for the same values of ω_n and ζ the rate at which the harmonic response falls off with rising frequency is greatly reduced. The

acceleration lag coefficient is given by $\frac{1}{\omega_n^2}$ compared with

$\frac{1}{\omega_n^2} \cdot (1-4\zeta^2)$ for the simple error control servo. We shall compare

the performance of this second order servo having zero velocity lag with that of a first order servo under the following conditions :-

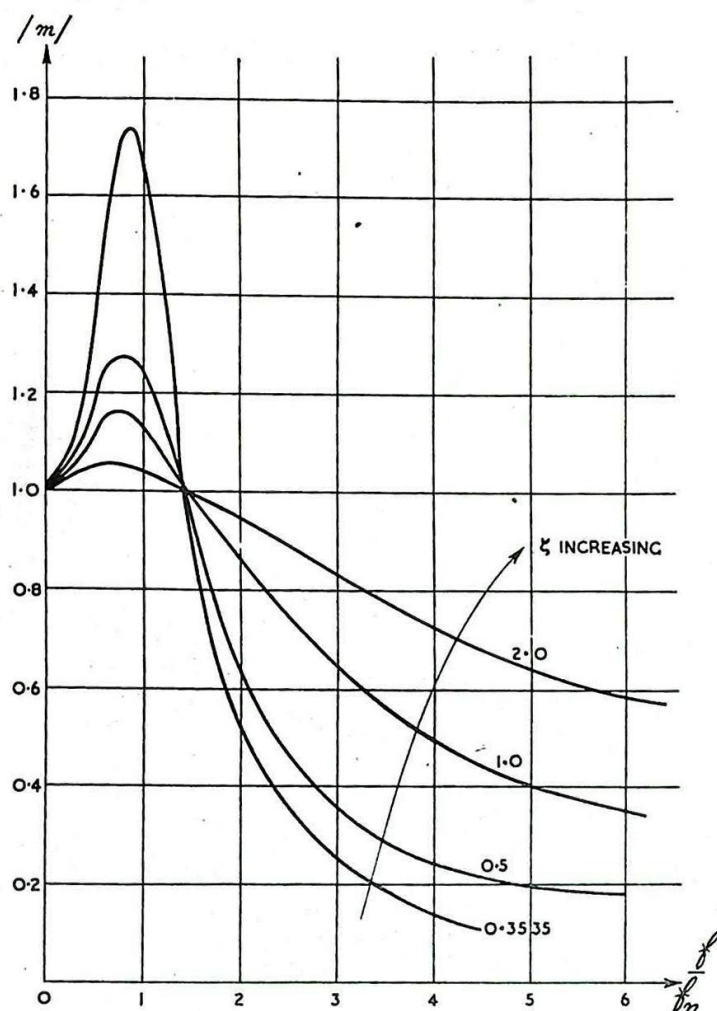


Fig. 887 - Harmonic response of second order servo (zero velocity lag).

- (i) Maximum angular velocity of the input = $10^\circ/\text{sec}$.
- (ii) Maximum angular acceleration of the input = $1^\circ/\text{sec}^2$.
- (iii) At 2 c/s the response must fall to 0.4 of its value at $f = 0$.

For the first order servo, putting $\omega = 2\pi f = 4\pi$ when $f = 2$ c/s, and $|m| = 0.4$ in equation (8), we obtain

$$a = \frac{1}{4\pi} \sqrt{\frac{21}{4}} = \frac{1}{12}.$$

Since the velocity lag coefficient is a , the maximum velocity lag is

$$10 \cdot \frac{1}{12} = 0.8^\circ.$$

For this servo acceleration lag may be neglected.

The acceleration lag coefficient for the second order servo is $\frac{1}{\omega_n^2}$; hence, with the same permissible error, we obtain

$$1 \cdot \frac{1}{\omega_n^2} = \frac{10}{12}$$

$$\text{so that } \omega_n = \sqrt{1.2} \doteq 1.1.$$

$$\text{Hence, at 2 o/s, } \frac{f}{f_n} = \frac{\omega}{\omega_n} = \frac{2 \cdot 2\pi}{1.1}$$

$$\doteq 12.$$

Fig. 887 does not permit of extrapolation to this extent, but calculation gives the result that for $\frac{f}{f_n} \doteq 12$ and $|m| = 0.4$,

$$\zeta \doteq 2.6.$$

Both servos then satisfy the given conditions. They may now be compared for pull-in time in response to unit-function input.

For the first order servo the time which elapses before the error is reduced to 10% of the input is given by

$$t = \frac{a}{0.4343}$$

$$= \frac{1}{12 \cdot 0.4343}$$

$$\doteq 0.2 \text{ secs.}$$

For the second order servo the time which elapses before the error is zero (there is a subsequent overshoot) may be shown to be given by

$$\epsilon^{(2\sqrt{\zeta^2-1})t} = 2\zeta^2 - 1 + 2\zeta\sqrt{\zeta^2-1}; \text{ (provided } \zeta > 1\text{).}$$

Putting $\zeta = 2.6$, we may deduce that

$$\epsilon^{2.2 \cdot 4t} = 25$$

$$\text{or } t = \frac{1}{4.8} \cdot \log_e 25$$

$$\doteq 0.67 \text{ secs.}$$

It is possible to reduce the time to the first zero by choosing a smaller value of ζ , and this permits of a larger value of ω_n . If we make $\zeta = 1$ condition (iii) is satisfied provided

$$\frac{f}{f_n} \geq 5, \text{ so that } \omega_n \leq \frac{4\pi}{5}.$$

Taking the maximum value of ω_n we substitute this value in the abscissa of the appropriate curve in Fig. 886 at the point where $\theta_0 = 1$,

$$\text{i.e. } \omega_n t = 1$$

$$\text{so that } t = \frac{5}{4\pi} \approx 0.4 \text{ secs.}$$

Any further decrease in ζ leads to excessive overshoots so that the time to the first zero error is not an adequate criterion.

In general the second order servo is to be preferred because of the smaller lags which may be obtained for a given frequency band-width, although where rapidity of pull-in is of primary importance the first order servo may be better. However, it is not possible to satisfy condition (iii) with the first order servo and at the same time reduce the angular lag to much less than one degree (in the instance considered), and this limitation may be prohibitive.

POWER SUPPLIES

1. INTRODUCTION

Two main sources of power are generally required for operating thermionic valve circuits. They are (i) Cathode heating power, and (ii) Anode power. Biasing voltages are usually obtained either from (ii) or from a separate circuit similar to the one used for (ii).

(i) Cathode heating power

In radar circuits the cathode power is almost always derived from small heater windings on the input power transformers. It could also be provided by separate AC or DC generators, or by storage batteries.

(ii) Anode power

This may be supplied by

- (a) Rectifier-filter circuits, operating from external mains or from motor alternators.
- (b) DC generators.
- (c) Vibrator packs operating from low tension storage batteries.
- (d) High tension batteries.

(c) and (d) are normally used only for very low power purposes and have little application in Radar equipments save in small short-range portable sets or test equipment. For radar purposes (a) is nearly always used.

The first few sections of the chapter deal briefly with rectifier-filter circuits. A fuller account is given in Admiralty Handbook of Wireless Telegraphy Vol. II. Sec. H, and other standard works.

RECTIFIER CIRCUITS

2. General

Either high vacuum tubes or hot-cathode mercury vapour rectifiers may be employed. Mercury arc rectifiers are sometimes used where very heavy DC loads are taken, while copper-oxide rectifiers also have a limited application for small DC loads.

One of the chief advantages of the hard valve is that it can be built to withstand very high peak inverse voltages, yet still provide an anode current which is adequate for most purposes. The gas-filled valve, although more efficient and supplying greater currents than a hard valve of similar size, must not be subjected to large inverse voltages; other safeguards also are necessary, which make their use troublesome (see Chapter 6, Sec 37).

3. Single-phase circuits

A simple diode half-wave rectifier operating from a single-phase supply is shown in Fig. 888(a). Since the current i can flow only during alternate half-cycles, in the direction shown, the output voltage v_o

developed across the load is far from constant unless adequate smoothing is provided. The available DC power output is correspondingly low. Also, since current flows only in one direction through the secondary of the transformer, the core is permanently magnetised, and is thus more readily saturated. This circuit is seldom used except for high-voltage

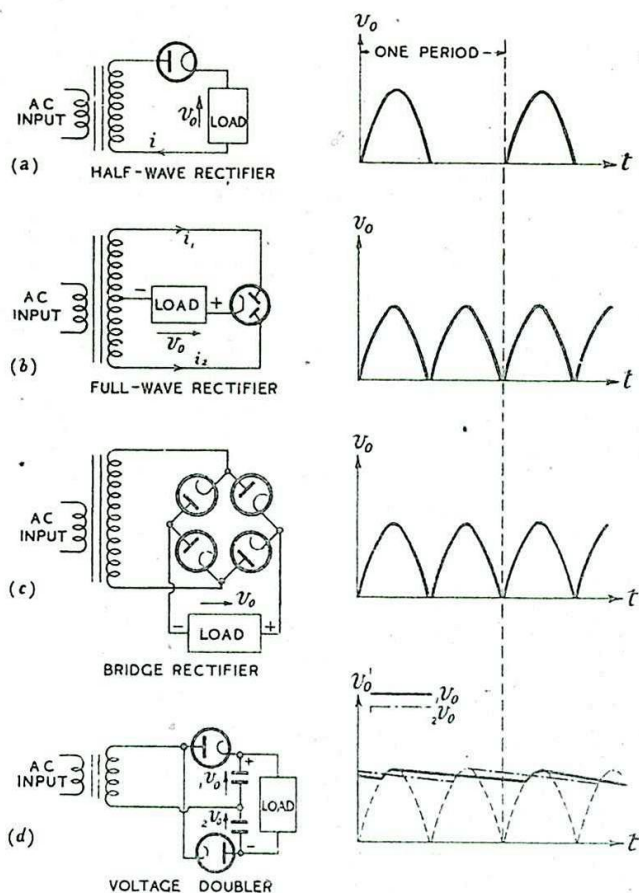


Fig.888.- Single-phase supply circuits.

low-current supplies such as are required for a cathode ray tube.

The most usual type of rectifier circuit is shown at (b), employing a centre-tapped secondary winding and either a double diode or two single diodes to give full-wave rectification. The ripple is at twice the input frequency, but the peak value of the output voltage v_o is only one half that obtained from a half-wave rectifier using the same transformer. v_o is, however, more constant and easier to filter, and since i_1 and i_2 flow in opposite directions through the transformer secondary, no permanent core magnetisation results.

A bridge full-wave rectifier is shown in Fig.888(c). This gives the same peak value of v_o for a given transformer as in (b), but four separate diodes are needed. Since the cathodes of these diodes are at greatly differing steady voltage levels, at least three separate heater supplies are necessary. This system is therefore most usually found in comparatively low-power circuits using four copper-oxide rectifiers instead of diodes.

Other circuits, such as the voltage doubler of Fig.888(d), are sometimes used for special purposes. So long as only a small load current is taken the voltage doubler gives a peak output voltage which is approximately twice that of the half-wave rectifier using the same transformer. Because of the large output impedance and high degree of

Fig. 889(c) shows a three-phase full-wave rectifier circuit in which each leg of the transformer secondary acts in the same way as the transformer secondary of Fig. 888(b). The output is the same as that obtained from the circuit of Fig. 889(b), but the transformer used is of simpler construction and no balance coil is necessary. The disadvantage is that four separate heater windings are needed on the filament transformer.

In general, the chief advantage of polyphase circuits is their superior output waveform, avoiding the use of intricate or bulky filter circuits. The main disadvantages are the more complex transformer systems and the large number of valves required.

FILTER CIRCUITS

5. General

The two principal types of filter circuit which will be considered are the series choke input filter and the shunt condenser input filter. C-R filters will be dealt with very briefly.

6. Series-choke input filter, (Fig.890(a))

The alternating component of the output voltage from the rectifier is developed chiefly across L_1 , whereas almost all of the steady voltage component is developed across C_1 . Further smoothing is provided by L_2 and C_2 .

The main disadvantage of this method is that the output voltage is considerably smaller than the peak input to the filter, due to the voltage drop across L_1 . The advantages are that the output voltage is steady, substantially independent of load current, and that the peak current supplied by the rectifier is not very large.

This type of filter is therefore most commonly used where good regulation is necessary because of varying load, and where a very high voltage is not needed.

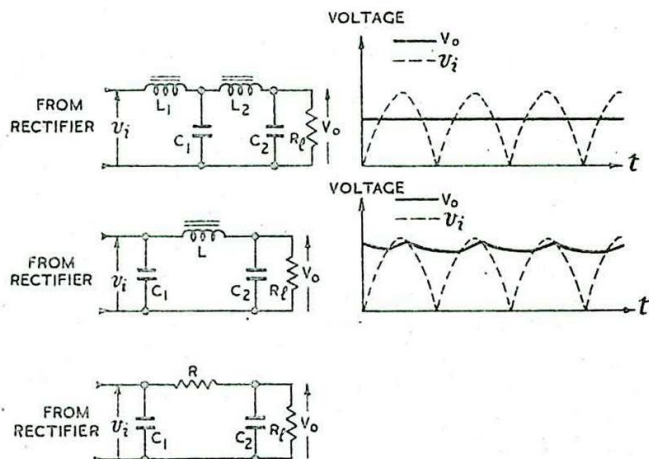


Fig.890.- Filter circuits.

7. Shunt-Condenser Input Filter (Fig.890(b))

In this type of filter the input condenser C_1 (because of its low impedance), charges very rapidly almost to the full peak voltage of the rectifier supply. This condenser voltage is very susceptible to changes in loading, however much smoothing is provided by L and C_2 . If R_L is substantially reduced the magnitude of the ripple across C_1 increases considerably, and the final voltage delivered to the load is much lower. In addition, the peak current supplied by the rectifier is much larger than in the circuit of Fig. 890(a) so that a larger rectifying valve is needed.

of smoothing necessary its chief use is for supplying power to cathode ray tubes, or other high resistance loads, where long time-constant R-C filters are practicable.

4. Polyphase Circuits

As an example of the advantages and disadvantages of polyphase circuits, three-phase supplies will be discussed.

Of the enormous number of possible three-phase circuits only a few are commonly used, the three most important being shown in Fig. 889. A three-phase half-wave rectifier circuit is shown at (a), in which each leg of the three-phase transformer acts in the same way as the single secondary shown in Fig. 888(a). The ripple frequency is thus three times that of the input, and its amplitude is much less than in the single-phase case, so that filtering is simpler. A three-phase transformer is used instead of three single-phase transformers in order to avoid DC saturation, following permanent magnetisation of the cores.

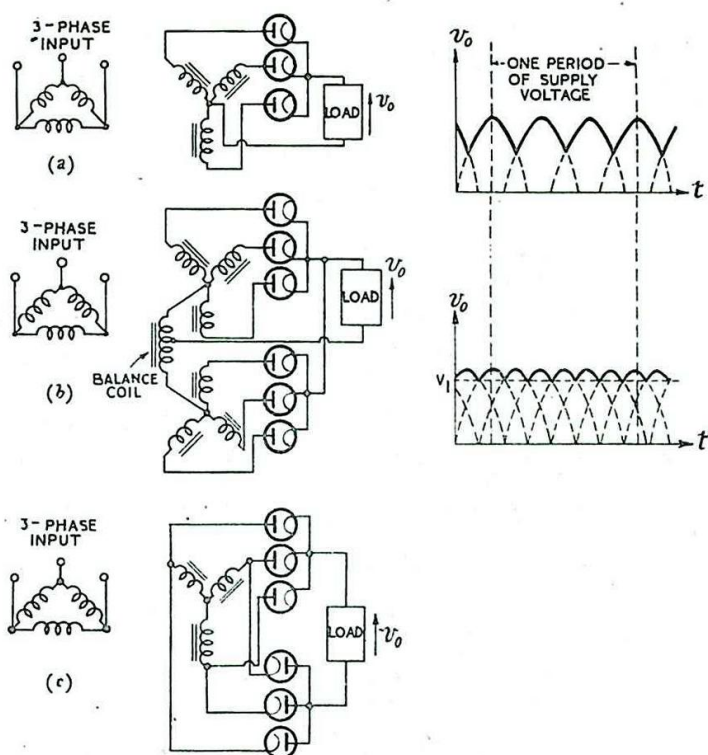


Fig. 889.- Three-phase rectifier circuits.

The three-phase half-wave double-Y rectifier circuit shown at (b) consists of two circuits of the type shown at (a) connected in parallel with the load in such a way that when the output voltage of one three-phase unit is a maximum, that of the other is a minimum. The amplitude of the ripple is smaller than that for the circuit of Fig. 889(a) and its frequency is six times that of the input. The balance coil which connects the two units in effect enables each to operate independently of the other, so that one valve of each unit can conduct simultaneously. With a resistive load and no smoothing, two valves are then conducting all the time so that each valve supplies current for one-third of the operating time. Without the balance coil each valve would cease to conduct when its individual supply voltage fell below the level V_1 , and would therefore supply current for only one-sixth of the operating time.

In three-phase circuits the main advantage of the shunt-condenser input filter is not nearly so pronounced because of the smaller ripple voltage, so that the choke input is much more largely used in polyphase circuits. In single-phase circuits the shunt-condenser input is more common, and if better regulation is required, some stabilising or regulating device is incorporated.

8. C-R Filter (Fig. 890(c))

Where the load impedance is very high, of the order of megohms, it becomes practicable to use C-R rather than L-C smoothing circuits, since the power dissipated in R (Fig. 890(c)) can be made negligible. The behaviour of this circuit is similar to that of the condenser-input filter described above, and it suffers from the same disadvantages. The amount of smoothing provided depends on the time constant CR, so that R should be chosen as large as possible without causing too severe a drop in output voltage.

STABILISING OR REGULATING CIRCUITS

9. Requirements

Radar circuits frequently require that one or more of the following conditions is maintained.

- (i) The load voltage (or current) must be substantially independent of the fluctuations in amplitude of the generator supply.
- (ii) The load voltage (or current) must be substantially independent of variations in load impedance.

Of these, voltage regulation is the more common requirement; and, since current regulation is usually limited to high power filament supplies and is provided by patent regulated transformer devices, it will be dealt with only very briefly.

Most generators are fitted with some form of electro-mechanical voltage regulator, of the make-and-break contact or carbon-pile types, and no treatment of these is given here. We are concerned mainly with regulating devices which receive their input from a DC supply source and provide a DC output to the load. The behaviour of these devices is considered with regard to variations in (i) supply and (ii) loading.

Either neon stabilisers or hard-valve circuits may be used. The neon valve possesses both characteristics required for (i) and (ii), namely an inherent "reference voltage" to which the load voltage is stabilised, and a low Differential Output Resistance (i.e., the change in voltage for unit change in current) for voltages above this value. Hard-valve stabilisers are essentially servo systems, (Chapter 18), in which the low differential resistance is provided by direct voltage feedback. There must be a reference voltage with which to compare the load voltage, and either a battery or a neon valve is usable, the latter being more common.

Instead of a valve being used a non-thermionic stabiliser may be employed. Examples of this are given in the thermistor circuits described below.

10. Neon Valve Stabilisers

The neon-valve-resistance combination shown in Fig. 891(a) is the simplest form of voltage stabiliser. Various loadlines are shown at (b), superimposed on the neon valve current-voltage characteristics, for different values of resistance and supply voltage. The anode voltage remains at the extinction voltage V_E provided:-

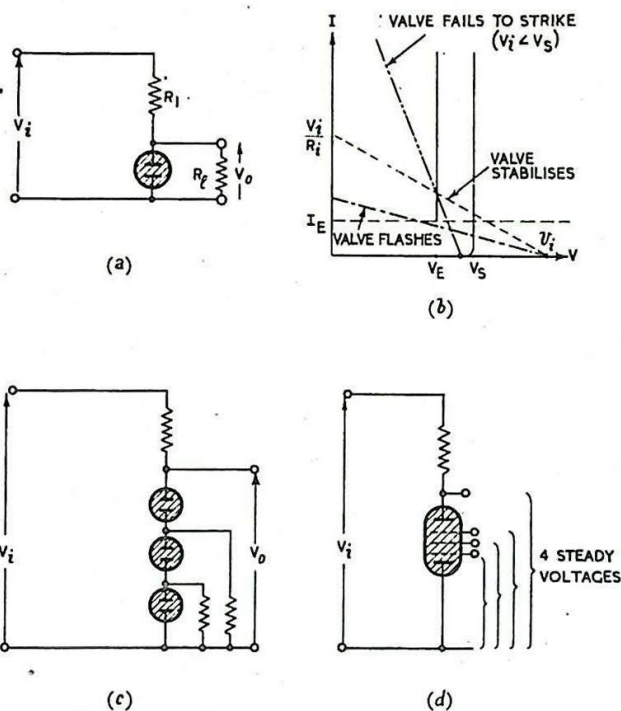


Fig.891.- Neon valve-resistor stabilising circuits.

- (i) The supply voltage is sufficiently high to cause the valve to strike, and
- (ii) The load resistance is sufficiently small so that the valve current is not reduced below its extinction value (I_E). If this happens the valve ceases to conduct, the anode voltage rises to the striking value, at which the valve reconducts and the familiar "flashing" of the neon valve occurs. Under these conditions the output voltage is anything but constant.

For a given load R_L and supply voltage V_i there are limits between which R_1 must lie in order to satisfy both of these requirements. A substantial change in either R_L or V_i may mean that one of the requirements is no longer fulfilled and the valve ceases to stabilise.

Within the working range determined by the chosen value of R_1 , the neon valve satisfies both of the fundamental requirements of Sec. 9, but the range of operation is restricted by the invariability of the striking and extinction voltages.

For supplies of several hundreds of volts three or four neon valves may be used in series. To prevent instability due to one or more of the valves flashing, it is usual to shunt each valve with an extra resistance as shown at (c).

As was shown in the case of the resistive anode load R_1 , the neon valve ceases to stabilise if R_1 is greater than or less than certain limiting values. If R_1 is replaced by another neon valve or by a neon valve in series with R_1 , there is no longer any limit to the magnitude of the anode load, and continuous relaxation oscillations can arise between the two valves. If a suitable resistance is shunted across each of the valves it will act like the load resistance R_L of Fig. 891(a) and will also ensure that the other neons are not extinguished.

Instead of several valves in series, a stabilivolt, or multi-electrode neon, may be used. This contains, besides the anode and cold cathode, several (usually three) intermediate gridded electrodes, as shown at (d). The whole envelope is filled with neon gas at low pressure, and the effect is similar to that of having four separate neon valves in series, with the difference that the gas in all four sections must ionise (and de-ionise) at the same time. The stabilivolt is generally made with all five electrodes brought out as terminals so that the intermediate values, as well as the full output, can be used. As in the case of separate neon valves, each unused electrode should be connected to earth through a suitable resistance.

11. Thermistor Stabilisers

The characteristics of thermistors are described in Chapter 6 Sec. 45, and Fig. 318 shows a typical current-voltage characteristic for one of these elements. Over a wide range of values of current the differential resistance is negative. If we connect in series with such a quasi-conductor a positive resistance equal to the negative differential resistance of the thermistor the voltage drop across the two is approximately constant; this is illustrated in Fig. 892(a).

Such an arrangement is usable in much the same way as a neon valve to stabilise a load voltage, as indicated at (b). Although the regulation is not nearly so good, in the simple circuit, as that provided by a neon valve, the element is much more robust, and is not affected, as is the valve, by limitations of striking and extinction voltages. Furthermore, by the use of more complicated bridge circuits, it is possible to reduce the load voltage variations to one part in a thousand over a wide range of load currents.

Thermistors are affected by changes in ambient temperature, but it is possible to compensate for this either by the use of thermostats or by employing additional elements with different temperature characteristics. Such a circuit is shown at (c), and the resultant stabilisation is illustrated by the two curves of Fig. 892(d), the voltage remaining between these values for all ambient temperatures between 60° and 110°F.

12. Hard-Valve Stabilisers

If a suitable constant reference voltage V_c is available, with which the output voltage can be compared, hard-valve circuits provide a

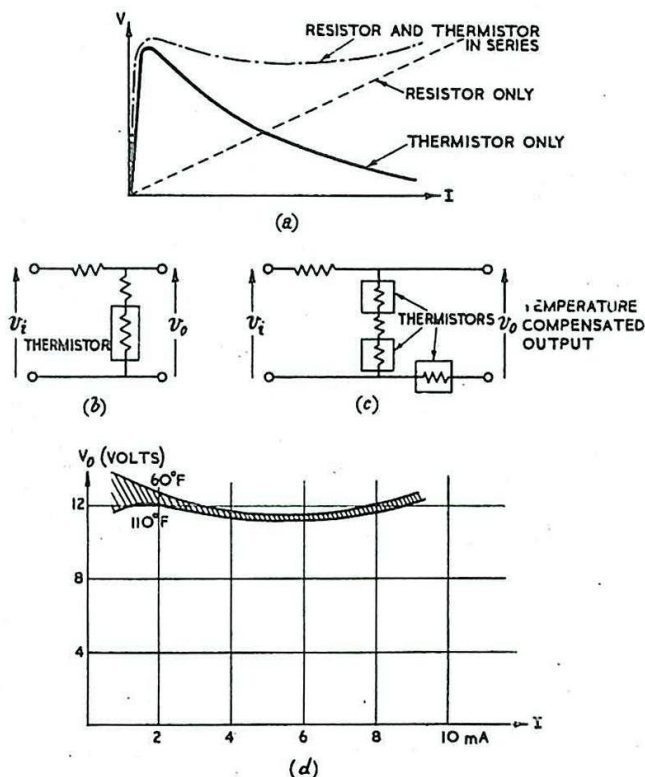


Fig. 892.- Thermistor stabilising circuits.

means whereby the load voltage can be maintained at a given multiple or fraction of V_c . This is a definite advantage over the simple neon-valve circuit, where the output voltage is pre-determined at the extinction value, or over the stabilivolt, where there are only four "spot" values from which to choose.

The fundamental servo circuit for maintaining a constant load voltage is shown in schematic form in Fig. 893(a). This circuit is readily adaptable to the stabilisation of load current, as illustrated at (b). If the voltage across a small (constant-valued) resistor in series with the load is maintained at a fixed level, the current through it, and through the load, is kept constant.

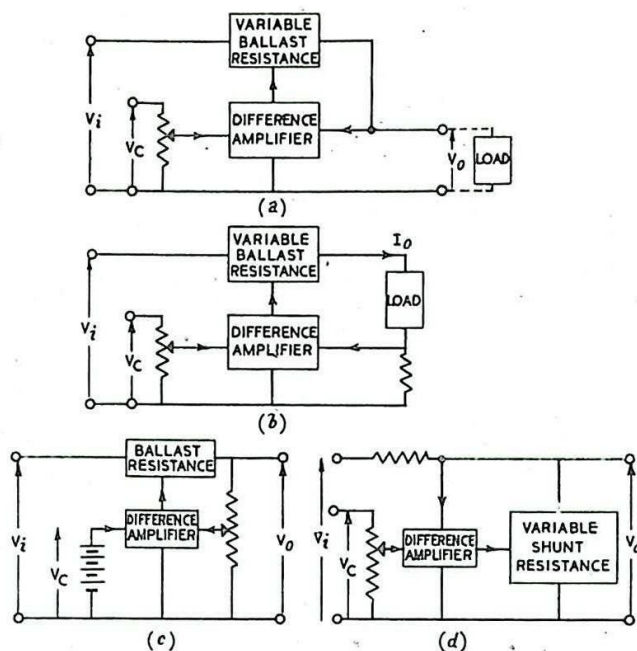


Fig. 893. - Hard-valve stabilisers; schematic arrangements.

In most cases the difference amplifier and the variable ballast resistance are thermionic devices, and the constant voltage is supplied by a neon valve fed from the variable supply. If the load voltage is greater than V_c , a reversal of the input networks to the difference amplifier, as shown at (c), will meet the case. By this means the neon may be replaced by a small constant-voltage battery.

An alternative arrangement to that shown at (a) is illustrated at (d). Here the ballast resistance is in parallel with the load, its magnitude being controlled by the difference amplifier, as before. This method is less economical than the former, the same power being taken from the supply for all values of load.

An arrangement corresponding to the schematic circuit of Fig. 893(a) is shown in Fig. 894(a). Two triodes, valves 1 and 2, fulfil the functions of difference amplifier and ballast resistance, respectively.

If the load voltage tends to rise, due either to a change in load resistance, or to an increase in supply voltage, the grid potential of valve 1 rises and its anode potential falls. This reduces the current in valve 2, thus increasing its resistance and causing a fall in the load voltage which partially offsets the initial rise.

An improvement may be effected by using a pentode in place of the triode as valve 1, the screen supply being taken from the input side of valve 2. A rise in input voltage increases the current in valve 1 and lowers the anode voltage as before, providing an additional stabilising effect. An additional advantage is the higher amplification provided by the pentode. The sensitivity of the stabiliser may be controlled by the insertion of a variable feedback resistor in the cathode load of valve 1. The larger this resistance is made without causing valve 1 to work on the curved portion of its characteristics the less sensitive is the regulation, but the larger is the range of voltage fluctuations over which the regulator will operate.

It is possible to dispense with the difference amplifier, but only at the expense of efficient regulation. The cathode follower circuit of Fig. 894(b) does this, and combines a fair degree of stabilisation with economy of material. A similar economy is provided by the direct voltage feedback circuit shown at (c). These circuits correspond to those of Fig. 893(a) and (c) respectively.

More elaborate, and more effective, circuits are illustrated by Figs. 894(d) and (e). At (d) the current through R_1 is governed by the control grid potential of valve 1, at which the full variations of the load voltage appear. These are offset by the corresponding variations of the resistance of valve 2. The screen voltage of valve 1 may be assumed to have little regulating effect.

By moving the slider of R_3 the constant output voltage may be varied within wide limits (raising the slider lowers the output voltage).

If extra resistors R_4 and R_5 are added as at (e), the output impedance of the regulator may be reduced to zero, or even made negative, so that an increase in load current causes an increase in load voltage. It can be shown that provided the current taken by the shunt valve circuits is small compared with the load current, this output resistance is approximately equal to $\frac{2R_4}{k} - \frac{R_5 R_4}{R_2}$, where k is the attenuation,

defined as the ratio of the input voltage variation to the output voltage variation. This attenuation is given approximately by

$$k = \frac{10^{-2} \mu R_1 R_2}{R_2 + R_5},$$

assuming valve 1 is a high gain pentode.

By a suitable choice of valves and components k can be made of the order of a million.

In practice, in each of the circuits of Fig. 894, it may be necessary to use several valves in parallel in place of valve 2 to carry the full load current.

REGULATED TRANSFORMERS AND AC STABILISATION

13. General

Where the stabilised output required is an alternating one, the circuits described in Secs. 9 to 12 are inapplicable and some other

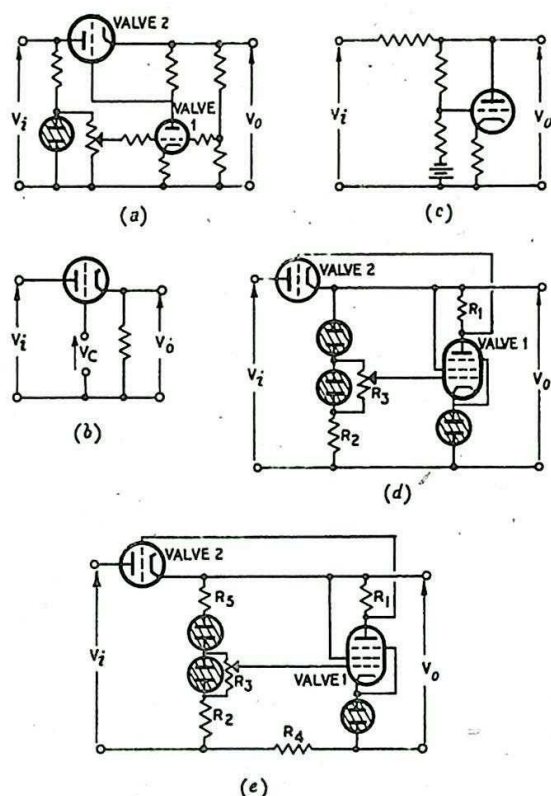


Fig. 894. - Hard-valve stabilisers: circuit diagrams.

device of the type described below must be employed. These devices are usually reliable and simple to operate but are not so adaptable as the thermionic ones. Their use of saturable iron cores gives rise to powerful extraneous magnetic fields; the power factor of the supply may be poor and the output waveform far from sinusoidal. These defects may be minimised by suitable screening, power-factor-correcting and filter circuits, but only with considerable increase in bulk, weight and cost. The frequency range over which these devices are effective is often very limited. They are not suitable for use with rectifier circuits for stabilising DC output because good regulation against input voltage fluctuations and against variations in loading cannot be successfully accomplished by the same regulating transformer.

14. Alternating Current Regulator

The essential circuit is shown in schematic form in Fig. 895. This is a simple servo in which the motor or other device moves the input tap of the regulating transformer so as to maintain the current in the secondary circuit at some constant level. This level may, for example, be fixed as the current necessary to close or open a spring contact which forms the difference element of the servo. It is usually an on-off switch, the control circuit when switched on driving the motor until the contact ceases to operate.

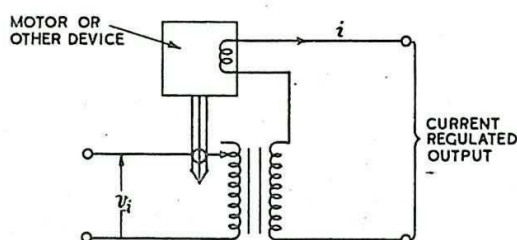


Fig. 895.- Current stabilisers.

In practice, the "motor or other device" may be a highly complicated system, depending upon the exact use to which the transformer is to be put.

15. Alternating Voltage Regulator: Motor Circuit

In this circuit, shown schematically in Fig. 896 the servo motor drives a tap on a large choke which shunts the AC input terminals. Between this tap and a fixed centre-tap is connected the primary of a transformer, the secondary of which is in series with the output line. This choke acts as an auto-transformer, reinforcing or reducing the output voltage according to the position of the variable tap, above or below the centre tap.

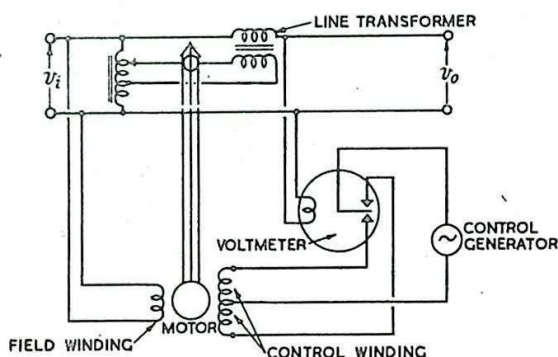


Fig. 896.- Motor regulator.

The AC servo motor may be driven by the field supplied by one of two control windings as shown. An AC voltmeter acts as difference element, and the servo is of the on-off type, the motor running at full speed in one direction or the other according to which of the two contacts is closed by the needle of the voltmeter. For smoother regulation multi-spring type contacts may be used, making the current in the control winding more smoothly variable.

16. Alternating Voltage Regulator: Saturable Reactor Circuit

Instead of the line transformer of Fig. 896, a saturable reactor may be used, avoiding the relatively complicated motor circuit. The voltmeter difference element is replaced by a difference transformer circuit using a thermistor, for producing a constant reference voltage from the variable supply.

The circuit is shown in schematic form in Fig. 897. The reactor consists of a small laminated iron core on which are wound two coils, the line and saturating windings. The more direct current there is supplied to the latter, the more the core tends to saturate and the lower the impedance of the line winding. It is on this variation in output impedance that the regulation of the transformer depends.

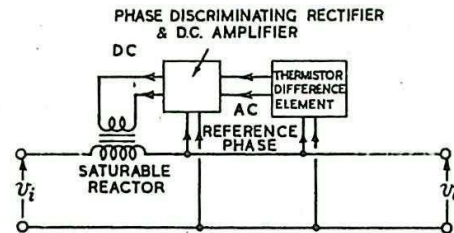


Fig.897.- Saturable reactor regulator.

The thermistor difference element is connected across the output terminals and the magnitude and phase of its output is indicative of the amount by which the line voltage is too high or too low (the "error"). This is rectified in a phase discriminating circuit (see Chapter 18 Sec. 14) so that the direct output current depends in magnitude and direction on the size and sense of the error. This current is applied to the saturating winding of the reactor.

There are several disadvantages in this elementary circuit, notably the loss in output power due to the absorption in the reactor, and the variable power factor of the output supply. These may be minimised either by using a separate supply for the reactor, the output from which is added to the line by a transformer in parallel with the line, or else by including the saturable reactor in the field circuit of the alternator instead of in its output line. The former method is indicated in Fig. 898. The latter is outside the scope of this chapter since it depends largely on the type of alternator field system used, and on other factors peculiar to the generator itself.

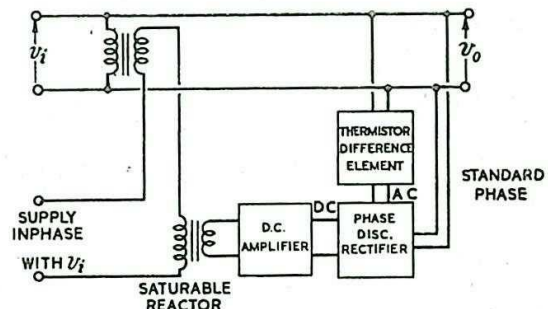


Fig.898.- Alternative circuit using saturable reactor.

The reactor-thermistor arrangements are applicable over a wide frequency range and are thus usable where motor-circuits are inefficient. The reactor introduces a certain unavoidable amount of waveform distortion, but not usually to the extent of affecting the equipment used in radar.

17. AUTOMATIC POTENTIOMETER (ELECTRONIC POTENTIOMETER) EARTHING CIRCUIT.

In some radar equipments it is necessary to use a power supply which is not connected to earth at either the negative or the positive

line or "rail". Generally, in such a case, it is sufficient to connect a potentiometer across the supply and to earth some point on this potentiometer. Provided the resistances between this point and the positive and negative lines remain constant and the supply voltage is unchanged, the lines remain at fixed potentials. If, however, leakage develops between either the positive or the negative supply line and earth, the resistance between this line and earth is reduced and the potential levels are upset; (Fig. 899). This is quite likely to occur with high voltage supplies. The effect may be minimised by making the resistance of the potentiometer as small as possible, but heat losses limit the practicability of this method. What is required is a device for maintaining the chosen point on the potentiometer sufficiently close to earth potential without actually connecting it to earth. The leakage currents which flow from the positive or negative lines to earth do not then flow through the potentiometer, and the resistances of the two portions, and therefore the potentials of the supply lines, remain constant.

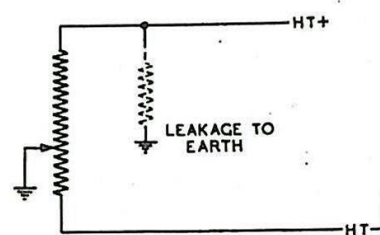


Fig. 899. - Direct earthing of the supply.

A device for producing the desired effect is the Automatic Potentiometer Circuit, or Electronic Potentiometer. A simple form of this circuit is shown in Fig. 900 . A cathode follower circuit is connected across the supply. Its control grid is connected to the required point on the potentiometer and the cathode is connected to earth. Large leakage currents may flow from either of the HT lines to earth without seriously affecting the action of the cathode follower, the grid-cathode voltage of which remains substantially constant. Hence the supply lines remain at reasonably constant potential levels with respect to earth.

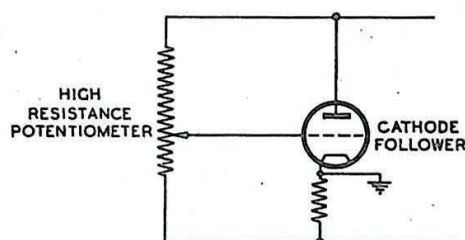


Fig. 900. - Use of cathode follower.

If the cathode load resistor is connected to earth at a point a few volts below the cathode the grid potential can be brought nearer to earth without the danger of grid current flowing and introducing a form of leakage which would produce the very effect which it is the object of the circuit to avoid.

An improvement in the simple arrangement can be affected by the substitution of a constant-current device for the cathode resistor. Provided the cathode follower valve is a pentode, as shown in Fig. 901 , variations of valve current with grid-cathode voltage are relatively independent of anode-cathode voltage. Hence, if the load is truly a constant-current device no variation in grid

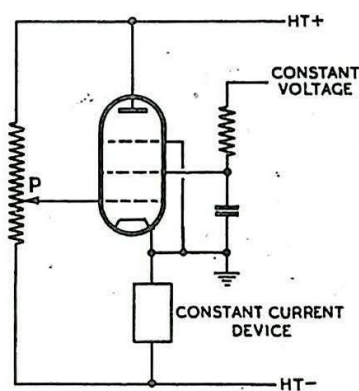


Fig. 901. - Use of constant current load.

voltage occurs however the position of the potentiometer tap P is varied. Leakage currents, however, still cause the grid-cathode voltage to vary.

A practical variant of the circuit of Fig. 901 is shown in Fig. 902. The lower valve is a Constant-Current Pentode, i.e., a pentode with the potentials at control and screen grids fixed with respect to cathode potential so that the valve operates above the knee of its $I_a - V_a$ characteristic. The screen grid of the upper valve is at a fixed potential relative to earth.

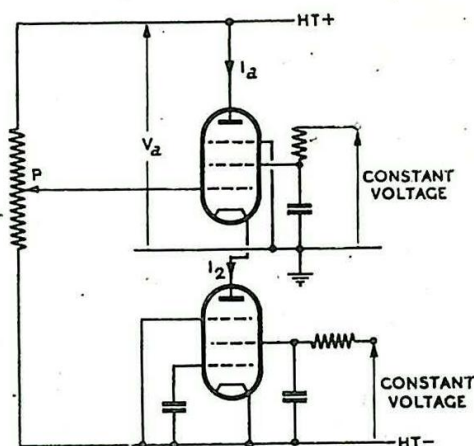


Fig. 902. - Use of constant current pentode as load.

Let I_a denote the anode current of the cathode follower valve. In the absence of leakage this is also the current through the load. If a leakage current I_+ occurs from the positive line to earth the current I_2 through the load (lower valve) becomes $I_a + I_+$, whereas if a leakage current I_- flows from earth to the negative line,

$$I_2 = I_a - I_-$$

If the leakage is due to a resistor R_+ in the first case, I_+ is given by

$$I_+ = \frac{V_a}{R_+}; \text{ whereas in the second case}$$

$$I_- = \frac{V_B - V_a}{R_-}$$

Fig. 903 shows the effect of leakage in a practical case in which it is necessary for P to vary over a wide range. If there is no leakage (curve A) the variation in grid-cathode voltage as V_a is varied from 1000 to 2400 volts, corresponding to different positions of the potentiometer tap P, is 0.3 volts. If there is a leakage resistance R_+ from the positive line to earth the current I_a is given by

$$\begin{aligned} I_a &= I_2 - I_+ \\ &= I_2 - \frac{V_a}{R_+} \end{aligned}$$

This is shown at curve B for $R_+ = 20 \text{ Mn}$. Similarly if there is a leakage resistance R_- from the negative line to earth the current I_a is given by

$$\begin{aligned} I_a &= I_2 + I_- \\ &= I_2 + \frac{V_B - V_a}{R_-} \end{aligned}$$

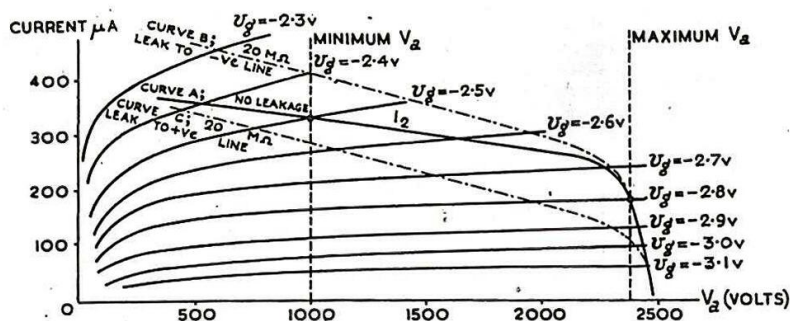


Fig. 903.- Effect of leakage.

This is shown at curve C for $R = 20 \text{ M}\Omega$. In neither case does the grid-cathode voltage variation over the required anode voltage range exceed 0.5 volts. Hence, whatever the position of P within the required range, the supply lines are always within 0.5 volts of their required values.

We may compare the above results with those which would be obtained by the arrangement of Fig. 899, for a potentiometer causing the same drain on the supply as the electronic potentiometer. The resistance of this potentiometer would be $\frac{2500}{300} \cdot 10^6 \Omega$; i.e., approximately 8 M Ω .

Without leakage the appropriate earthing point would be $\frac{1000}{2500} \times 8 \text{ M}\Omega = 3.2 \text{ M}\Omega$ from the positive line for $V_a = 1000$ volts.

With 20 M Ω leakage resistance from the positive line to earth the 3.2 M Ω is modified to become

$$\frac{3.2 \cdot 20}{23.2} = 2.76 \text{ M}\Omega.$$

The 2500 volts are therefore distributed about earth potential in the ratio

$$\frac{2.76}{4.8} \text{ instead of } \frac{3.2}{4.8},$$

so that the positive line is 913 volts above earth instead of 1000 volts.

Hence a shift of nearly 100 volts occurs in the potentials of the supply lines, compared with 0.5V in the electronic potentiometer circuit.

